



Commentary

Comments on the “SSF Report” from the perspective of economic statistics

Dong Qiu¹ and Dongju Li^{2,*}

¹ School of Statistics, Jiangxi University of Finance and Economics, Nanchang, 330013, China

² School of Statistics and Big Data, Henan University of Economics and Law, Zhengzhou, 450046, China

* **Correspondence:** Email: lidongju0223@126.com.

Abstract: Sustainable development measurement is an important exploration field in socio-economic statistics, which has been attached great importance by society for 30 years. In 2010, the Commission on the Measurement of Economic Performance and Social Progress (CMEPSP), headed by J.E. Stiglitz, A. Sen and J.P. Fitoussi, published a report entitled “Mis-measuring Our Lives: Why GDP Doesn’t Add Up” (“SSF Report” for short). It systematically reviews and summarizes methods of economic measurement, of which classical GDP issues, well-being measurement and sustainable development measurement constitute the three main contents. Society is advancing and we cannot follow the existing measurement methods without carefully re-examination. This paper analyzes several measurement dilemmas hidden in GDP statistics, explores the feasibility and necessity of the well-being measurement, queries the sustainability of sustainable development measurement.

Keywords: GDP statistics; well-being measurement; sustainable development measurement; measuring dilemmas

JEL Codes: B23, I23

1. Introduction

In 2010, the Commission on the Measurement of Economic Performance and Social Progress (CMEPSP), headed by J.E. Stiglitz, A. Sen and J.P. Fitoussi, published a report entitled “Mis-measuring

Our Lives: Why GDP Doesn't Add Up". It systematically reviewed and summarized the methods of economic measurement, in which classical GDP issues, well-being measurement and sustainable development measurement constitute the three major contents.

The core message of the SSF Report is to change the focus of economic statistics from economic production to social well-being, which is a decisive reform in measurement perspective. The "Beyond GDP" agenda expands the measurement framework and pays more attention to people's quality of life (current well-being) and sustainable development (future well-being). GDP statistics are the traditional content of economic statistics and ought to be relatively mature in public perception. However, it cannot be assumed, without questioning and thinking, that GDP statistics are perfect or simple. In fact, there always remained a series of controversial issues in the previous revisions of the SNA. Only when the proposal of a disputed issue reaches a consensus, can it become part of the measurement standard. Otherwise, the dispute will be put on hold and unresolved issues still exist.

Focusing on the quality of life (economic wellbeing), this paper questions some ideas of the SSF Report and puts some critical thinking on the economic logic issues hidden in the process of the economic measurement from the two fundamental perspectives of feasibility and necessity. Adding more indicators is not the only solution to strengthen the current economic measurement. And it is also not a good choice to establish a new model of wellbeing measurement to replace the GDP system. A "GDP plus" model seems to be a better option.

Further, sustainable development (future well-being) is an important part of the SSF Report. This paper includes the reflection on the following four issues: the concept premise of sustainable development; the concept premise of sustainable development measurement; four methods of sustainable development measurement; the methodology of sustainable development measurement.

Finally, in 2018, the OECD released a new economic measurement report (the author referred to as the SFD Report). The SFD Report is a follow-up study of the SSF Report. After the global economic recession, the economic downturn and income inequality have become more prominent, making the original SSF Report too general. Therefore, the main contents of SFD Report and its comparison with the SSF Report are analyzed.

2. Main text

2.1. Analysis of several measurement dilemmas implicated in GDP statistics

This section analyzes four indicators in GDP statistics: Actual Final Consumption (AFC); Government Output; Unpaid Housework; Defensive Expenditures.

2.1.1. Measurement of actual final consumption

In economic measurement and national accounting, there is a moderate coordination issue on the measurement boundary of indicator setting and in-depth analysis. To what extent should the indicators be refined? How to deal with the relevance issues caused by the refinement of indicators? We take AFC as an example to illustrate.

Necessity of measuring AFC

SNA1993 distinguishes Household Final Consumption Expenditure (FCE) and Actual Final Consumption (AFC) as:

$$\text{Actual Final Consumption} = \text{Final Consumption Expenditure} + \text{Transfers in Kind}$$

The motivation for further measuring AFC lies in that different countries have different government arrangements for households' welfare systems. The governments of developed countries provide more public goods for households, so households pay much less for the final consumption compared to their actual consumption levels, resulting in the inconsistency between the two final consumption states. According to the Institutional Irrelevance Principle (Qiu, 2019a), economic measurement should not be affected by different institutional arrangements. For example, for measuring Rent in SNA, whether it is rental or self-owned housing, all should be included in the accounting scope. Rent is calculated based on the actual rent. In addition to FCE, the setting of AFC also implies the idea of measuring economic results that have nothing to do with institutional arrangements.

Different countries have different levels of final consumption. The level of direct public services in Europe is higher than that in the United States in terms of the government's provision of education and medical care for households. In this sense, the well-being level of European households in terms of their FCE is far lower than that of the United States. But if we look at AFC, the gap of welfare between them is not that big. Therefore, distinguishing these two consumption indicators facilitates a clearer representation of the welfare level and the structure of different countries.

From a political point of view, with this indicator distinction, European governments could show the voters that their governance efficiency is not worse than that of the United States, at least not that different. The enlightenment of this case for economic measurement is that economic facts are essentially a huge and complex existence, so that the measurers can observe, measure and calculate from different angles and in different patterns. The setting and selection of indicators have their socio-economic background, and the measurers always have their economic standpoints. The same final consumption is an economic fact when viewed from the perspective of Household Expenditures, and another economic fact when viewed from the perspective of Actual Consumption. Which facts to adopt on what occasions should depend on the demand for analysis.

In daily transactions, households provide labor to get Benefit in Cash, and get goods and services in kinds that they need through Payment in Cash. However, in many cases, households directly receive Benefit in Kind (BiK). The in-kind benefits provided by the government to households can be attributed to Transfer in Kind (TiK). The company may provide in-kind benefits in addition to the salary, which is in fact a Return in Kind to employees' labor. For employers, this is Payment in Kind (PiK). The government provides households with free education and medical services. This kind of Payment/Benefit in Service (PiS/BiS) is also a non-cash way, which is quasi-in kind. With the help of the distinction between in-cash and in-kind items, we seem to be able to analyze the difference between FCE and AFC more clearly.

How many forms of payment or benefit exist?

The problem is that while this distinction between in-cash and in-kind brings measurement benefits, it also implies the uncertainty problem caused by the different levels of refinement of the relationship between indicators, opening the paradox for economic analysis.

Regardless of benefits or payment, in-cash and in-kind do not constitute the entire set of cost (or benefits) concept. In addition to in-kind sub-items, non-cash items also include other sub-items. We cannot confuse Payment in Kind with non-cash payment. We should have a comprehensive understanding of households' compensation package, and correspondingly, we should also consider the payment package of employers and consumer suppliers.

In addition to in-cash and in-kind, there are other ways of payment or benefits: Payment/Benefit in Service (PiS/BiS), Payment/Benefit in Information (PiI/BiI), Payment/Benefit in Consumption Environment (PiCE/BiCE), Payment/Benefit in Experience (PiE/BiE), Payment/Benefit in Opportunity (PiO/BiO), and perhaps others.

Observing the economic reality carefully, such cases abound. The company offers commuter cars, free hairdressing services, for free or only at internal prices. Personal information, such as home address, mobile phone number and email address, is required to get free delivery for products, to get free pictures taken at the airport, to apply for various membership cards and discounts. Such information can be used for other purposes and become part of commercial interests. For a caregiver for the elderly, in addition to salary, a life-long experience gained by communication with the elderly is an extra bonus, and for a tourist guide, it is the travel experience.

A prominent phenomenon in China is that many young people are willing to work in the four major international accounting firms, where income and effort are not proportional. Heavy workload makes people work overtime for days and nights. Why are young people still flocking to those firms? In fact, they don't intend to stay, but for three to five years of professional experience in the top firms in the field, which is a long-term, intangible, contingent benefit, so that they can jump to a higher position in another company and their income can rise several steps. From the perspective of employers, they know that they can provide such benefits that people long for, allowing them to pay salary lower than its labor intensity level. Because in addition to salary, high-end internship opportunities are also of great value. Even if the salary is not high, there is a constant flow of job seekers. Why is there such a market wage price? Both the supply and demand sides have considered the experience and opportunities, and the benefits along with, implicit in the position.

The distinction between in-cash and in-kind facilitates more accurate measurement and comparison of actual consumption levels and structures in different countries. However, according to this correct way of thinking, cash and non-cash forms should be distinguished, and non-cash forms should be further subdivided into in-kind, service, information, opportunities, experience and consumption environments.

According to the author's analysis, there are at least seven forms of payment or benefit, so:

$$\text{total} \frac{\text{payment}}{\text{benefit}} = \text{cash} \frac{\text{payment}}{\text{benefit}} + \text{non-cash payment/benefit}$$

$$\begin{aligned} \text{non-cash payment/benefit} = & \text{payment/benefit in kind} + \text{payment/benefit in service} \\ & + \text{payment/benefit in information} + \text{payment/benefit in opportunity} \\ & + \text{payment/benefit in experience} + \text{payment/benefit in consumption environment} \end{aligned}$$

Economic measurement dilemma caused by including TiK

If the above analysis is valid, the calculation formula of AFC equal to FCE plus TiK cannot be established. The relationship in SNA between TiK and income, consumption indicators is far from perfect. It's just a view of economic measurement limited to materialism, which some service

payments or benefits may be considered, but payments or benefits for information, experience and opportunities are not considered at all.

The problem is that if TiK should be included in the accounting, all non-cash payments (benefits) also should be included in the accounting. They constitute necessary components of AFC. The measurement logic is the same as that of in-kind and service. For households, information, experience, opportunities and consumption environments constitute payment or benefits like goods and services. If we want to measure all the actual consumption of households, it is necessary to take all non-cash forms of income (payment) into account, same as the case of TiK, otherwise the so-called AFC does not match its name.

In other words, basic accounting formula of SNA for households' income and consumption implies omissions:

Actual Final Consumption \neq *Households' Consumption Expenditure* + *Transfers in Kind*

Adjusted Disposable Income \neq *Disposable Income* + *Transfers in Kind*

TiK is only a sub-item of non-cash benefits (payments), which does not represent the remaining sub-items. We cannot even determine its proportion in all non-cash benefits. Only assuming the remaining sub-items of non-cash benefits to be zero, can the above accounting balance formula be established. In fact, the real accounting relationship should be:

$$\begin{aligned} \textit{Actual Final Consumption} \\ &= \textit{Households' Consumption Expenditure} + \textit{Total Non} \\ &\quad - \textit{cash Final Consumption} \end{aligned}$$

Adjusted Disposable Income = *Disposable Income* + *Total Non – cash Benefits*

It is valuable to really clarify the relationship between indicators, but meanwhile, it may open the Pandora's box and cause a series of difficult problems in economic measurement:

First, various economic activities overlap with each other. How to include benefits of services, information, experience, opportunities and consumption environment in the actual income (costs)? For example, what is the learning benefit of taking care of the elderly? What is the value of the non-cash enjoyment of accompanying tourists? How much is it worth working in the four major international accounting firms for a year of experience? Economic activities are diverse. This kind of non-cash benefits does not have a market price, nor corresponding or reference price like goods or even some services. They are often implied in transactions, with invisibility and contingency. How to include them in the estimation? How much influence will these factors have on the measurement output and price?

Second, now SNA only considers TiK, but ignores non-cash benefits. If we expand the accounting boundary, where should it end? Is it possible to include all the other categories that come to mind, or gradually include categories according to the feasibility of the measurement? Are there any other forms of non-cash benefits?

The major problem is that if the above forms of payments are established, the concept of unpaid work will be difficult to establish. No matter what activities people engage in, they can always get some form of non-cash payments or pay some non-cash costs.

What are the knock-on effects of removing unpaid housework on the measurement with major changes?

First, if unpaid work, one of the five basic activities listed in Time Use Account, no longer exists, should the popular Time Use Account in developed countries be redesigned? This is just the first Dominoes.

The author's argument also provides a measuring basis for unpaid housework to be included in GDP. But if unpaid work no longer exists, then the concept of Unemployment cannot exist. How can unemployment statistics be carried out? How to ensure employment? This is the second.

In addition, will the concept of Defensive Expenditures still exist? Economic principles tell us that there is no free lunch in the world. In fact, the inherent logic of this axiom can only be clarified only when all forms of benefits are fully considered. If we are convinced that everything has a cost, then correspondingly, no activities are free. The difference may be a structural issue of its utility, whether direct or indirect, potential or apparent, large or small, long-term or short-term, etc. In any case, there will always be some forms of benefits on payments. As far as a certain space and time is concerned, it is difficult to determine what the utility is, and how to talk about "defensive"?

In a word, once the item of TiK is included in the economic measurement and accounting, other non-cash benefits have corresponding jurisprudence of existence, the measurement of unpaid housework and even Time Use Accounts, Unemployment, and Defensive Expenditures will be the most at risk.

So, to what extent should the indicators be refined? This involves the Production Boundary Principle from the internal perspective (Qiu, 2019a). It must also refer to the Principle of Moderate Coordination of Measuring Boundary (Qiu, 2019a). Different economic measures cannot fundamentally contradict each other. The improvement of one economic measure should not jeopardize other measures, just as including unpaid housework into GDP will make Unemployment no longer exist.

2.1.2. Measurement of government output

Supplementary calculation of Government Output

In SNA, Government Output is measured by its expenditure. Although Government Expenditure can reflect its economic role, it is not complete and measures only part of its output. Consideration should also be given to government policy service as a part of Government Output, whose main function is to smooth the economic cycle (Qiu, 2019b).

Alan Greenspan pays more attention to the social role of GDP function: GDP has brought order to an otherwise chaotic world. This statement can support the author's view that government policy services should be regarded as Government Output. The economy is chaotic. One manifestation of chaos is radical ups and downs. With GDP, it is easy to grasp the trend of the economic cycle. The government can adjust policy based on this information to smooth out fluctuations and reduce potential losses.

However, this will bring another problem: if this part is included, the proportion of Government Output in GDP will be higher. This part must be only an estimate, because it is impossible to obtain an observation value. A measurement problem is that the estimated value in the whole GDP measurement is too large, which violates the Observed Value Priority Principle of GDP statistics. Moreover, the Government Output measurement also involves the identification of social pattern, which is usually the more developed the country, the more policy services and the better the quality. If this part of the government policy service output is added, can developed countries be described as "small governments and large societies"?

General and special expressions of composition of sector value added

Among the five major institutional sectors of SNA, consumers are mainly households and producers are corporations, the government collects taxes and provides public goods, while non-governmental organizations use charity fund to provide non-profit services, combining four types of institutional sectors related to the economy in foreign countries, in this way, space-time balance of the whole economy can be established.

The output measurement of corporation sector is relatively mature, and Value Added equals Factor Cost plus Profit (in fact, Profit is also a factor compensation), namely,

$$\text{Corporation Sector's Output (Value Added)} = \text{Factor Cost} + \text{Profit}$$

The government provides public goods and services in kind for society, which is a non-profit organization. Government Output is not easy to measure. If profit items that should not exist by the definition are removed, according to the method of Corporation Value Added, Government Value Added will remain Factor Cost. It seems logical to use Factor Cost to measure Government Output, that is to treat input as its output. But the author believes that a more general expression of output should be

$$\text{Output (Value Added)} = \text{Factor Cost} + \text{Benefit}$$

Profit is only a secondary concept, or derivative concept, of benefit in form of corporation, and benefit is the original concept. In the corporation sector, Revenue is mainly represented as Profit (of course, this is not the only case in the economic practice). Therefore, non-profit is not without seeking profits. Otherwise, what is the meaning of the existence of government sectors? To a large extent, government's participation in economic activities is to provide positive externalities to the society. It is not called the output of economic activities without seeking profits.

If the benefit is linked to the concept of externality, there can be

$$\text{Benefit} - \text{Profit} = \text{Externality}$$

When the externality is zero, Revenue equals Profit. Positive externalities exist, and Benefit outweighs Profit. Negative externalities exist, and Benefit is less than Profit. Usually, zero Profit is not equal to zero Benefit, only in the absence of externalities.

Positive externalities, as the output of government sectors, although not in the form of profit, should obviously be a more important part of Government Output. If only Factor Cost is considered, the output scale and level of government will be greatly underestimated.

It seems that there is a logical defect in applying the formula of Corporation Value Added to government sectors. It seems correct from the historical order of economic measurement. In fact, it is contrary from the logic of economic measurement. When determining Value Added of each sector, we should first consider the original principle, beginning with the concept of Benefit, which is expressed as Profit in the corporation sector, and what is the form of this Benefit expressed in other sectors? Further research is needed.

Efficiency of government output and social well-being

There is also an issue of positive and negative benefits caused by the level of public service. If the government policies often fluctuate, there will be a great negative benefit. Although subjective intention is to adjust counter-cyclically, if the timing is not right, this output will be a negative

benefit. How can a negative benefit be calculated as an output? In fact, in addition to physical production, other service outputs may also produce negative benefits. Therefore, the positivity or negativity of benefits cannot be used to determine whether to be included in the output. When we talk about the concept of benefit, the marginal benefit is either decreasing or negative. The direction and degree of benefit are random and cannot be used as a basis for calculating the output value.

Of course, it is difficult to measure the amount of government services. The question of whether unpaid housework should be calculated as output value will be discussed. Even if it should be measured, it is difficult. But difficulty cannot be an excuse for not measuring, and other items that are difficult to measure are measured anyway. In fact, there are no items that are easy to measure in economic measurement. It does not make sense that while other difficult-to-measure items are included, this item is neglected for its difficulty to measure, which is adopting different criteria to discriminate.

There is also the issue of efficiency of Government Output, which is a problem of cheap government. The government should not only be clean, but also cheap. The fact that the government does not pursue profit does not mean that it does not make profit, nor does it mean that it does not pursue efficiency. According to Factor Cost Substitution method, substituting input for output implies an assumption that the economic activities of this sector are not effective, and that output of government is its input, which does not make sense.

The development of developed countries lies in the high level of public services, which increases the well-being of the people. The quality of government output in developing countries is relatively low, the level of management is relatively low, and the increase in the well-being of the people is subject to many constraints. The SSF Report emphasizes that the better measurement of government-provided individual services is central to the better assessment of living standards. This shows that the measurement of Government Output is still an important basis for expanding from GDP statistics to well-being measurement.

Issue of measuring government value added by “direct method”

How to measure Added Value of government sectors? Developed countries now generally tend to adopt the output-based measure directly. Government public services, education and health care in developed countries account for the largest proportion of the two major items, but not in developing countries.

The output of education and health care was originally estimated using the input-based measure. To adopt the direct method, it needs to be clear first: what exactly is the output of medical and health care? Some hold a level view: output of medical and health care is nothing more than the level of people’s health. Some hold the incremental view: medical and health care should maintain and improve people’s health status, so the output of health care in a time period should be the improvement of health status, that is, health increment.

Special attention should be paid to the following issues in health measurement:

First, people’s health status is determined by many factors. Government-provided medical and health services will promote its improvement, but it also depends on the basic status of population health (pre-stock level), genes, diet, living habits, private medical investment and exercise methods.

Secondly, as far as human health status is concerned, not only are there many determinants, but they are not independent of each other. They will interact with each other, and their interactions may be multi-cycle and multi-level. If the variables are arbitrarily assumed to be independent in the measurement, it is easy to cause bias.

Third, people's health status has its own natural laws. After reaching the peak of health state, their physique will inevitably decline. Henceforth, slowing down the deterioration of health can be a positive effect of various factors. In other words, the interaction of various factors trying to improve people's health will inevitably be used to offset the trend of natural deterioration of physical fitness, but the compensation effect is different in different time and space. In short, in an aging society, even if the health status of people deteriorates, it may still be a positive result of the government's increase in medical and health services, otherwise people's health status should have been even worse. So, how should the degree of influence of this natural deterioration factor implied in the changing health status be considered in the measurement?

Fourth, there is time lag in the operation of government medical and health services. The increment of health improvement and the role of government investment in health care are not limited to the current measurement period. It is not that the government invests 10 billion yuan at the beginning of the year, people's health status can have a 10-billion-yuan worth of improvement in this year. The effect of medical and health input varies in different time periods. Some are first large and then small, while others are first small and then large, with different rate of change. It is similar to the estimation of consumption of Fixed Capital, and to the determination of its role, it requires at least two basic assumptions: One assumes the role (life) cycle of services invested, that is, how long will the impact of input last; and the other is the change rate of the force of each period. In order to estimate the increment of health improvement brought about by government medical and health investment in each period, we need to know the magnitude of its impact in different time periods, and the magnitude of the interaction of different factors and the superimposed effect of the time lag.

In summary, it is difficult to distinguish how much of the health level and its improvement (if there are increments) can be attributed to government medical and health services. It can't be done accurately, even approximately is not easy. Assume that the original health level is 1, now it is 1.8, how much of the increase of 0.8 can be attributed to government services? The results of multiple factors should be divided into various factors, which is a big issue. Can the existing methods of mathematical statistics do this dissection well? Perhaps, we can use the method of fuzzy comprehensive evaluation to explore. The author prefers to emphasize the fuzzy method, because the economic phenomena often encountered is that the occurrence of an event is determined, but the boundary state is not clear. In contrast, the fuzzy method may be better at dealing with this type of uncertainty.

The input-based method is only a rather rough method. Economists in developed countries are unwilling to use low-level measurement methods. Over the years, they have been trying to use the output-based method to measure output of government services. European Union countries and OECD countries are trying, but that is not to say that they can already do well.

One of measures to be explored is to adopt non-value output indicators, that is, physical indicators, in terms of education, such as the number of students trained or the cost of training per student and medical care aspects, such as the number of patients cured, and so on.

However, physical indicators cannot be aggregated. To reflect the macro state, further synthetic processing is necessary. How to record the quality change? If undifferentiated quantity measures, such as the global total number of students or patients, are adopted, changes in the composition of output and in its quality may be neglected. An important criterion for the reliability of output-based measures is that they are based on observations that are detailed enough to avoid mixing up the real change in volume with the compositional effect.

For example, if per-student fees increase, one might conclude that the unit cost of education has increased, but this may be the case. It may be due to the change to small classes or the increase in the number of students majoring in engineering. The small class costs more than the large class, and the output quality should be higher. Engineering requires more funds than liberal arts and sciences. The measurement mistake arises because the simple number of students is too undifferentiated for an output measure to be meaningful, so a more detailed structure is needed.

However, to what extent can it be called detailed enough? What is the quantity of the word “enough”? Is it possible to get the answer in advance? How far should the decomposition be? For example, is it enough for the class to be divided into only large and small classes? How many levels should it be divided into? Should we distinguish between different majors in engineering majors? Here, we should not only consider the necessity of measurement, but also consider the feasibility of measuring budget constraints. The OECD countries are at a relatively high level of market economy. Education and health care are well developed. Perhaps there are funds and professionals to do a detailed decomposition, but what about developing countries?

Countries with different levels of development face different measurement objects, different statistical capabilities and budget constraints. It is not easy to achieve consistent decomposition levels, and the quality of the overall data results will vary. What is the spatial comparability of the output data of government services with different quality in different countries? This is the case in ICP: rich countries use the direct method to measure educational output and other countries use the indirect method (input-based method) to measure educational output. Is the data inherently comparable to put the results of two different methods into one framework for international economic comparison? Are the results of forcible comparisons credible in quantity and quality?

In fact, it is because the direct method is unable to answer the above problems (of course, the understanding of these problems is not so clear so far), that the indirect method is adopted to replace the government’s public service output with input value. Only by solving these problems can we re-enable the direct method to calculate the output of public goods. Otherwise, in a certain sense, it is equivalent to returning to the starting point of measurement.

2.1.3. Measurement of unpaid housework

Housework is divided into paid and unpaid. Unpaid housework is also an economic activity and can bring economic benefits. Economic debate on this issue has reached a consensus. However, not all economic activities are included in Output Value, and the boundary of the two is not necessarily identical, which involves the coordination of different measurement principles.

The difference between unpaid housework and paid housework

Paid housework and unpaid housework are not exactly same in quality. Advocating that unpaid housework should also be included in Output Value includes the assumption that the cost-benefit comparison of the two is same.

Paid housework has been tested by the market, is up to standard, has job credit, and is a service that can be accepted by other families. Moreover, the degree of integration between the labor and the provider’s personal leisure is poor, and the practitioner is mainly for income.

Unpaid housework is not accepted by the market, and only has potential acceptance. Unpaid housework tends to be multi-purpose. It is more integrated with personal leisure and is also related to

the job characteristics of family members. Emotionally and rationally, in a family, it is often unemployed or flexible workers who undertake more housework. Although the market threshold for housework may be low, there could be many people who fail to meet the standard and be qualified as a houseworker. We should not think that there is no threshold for housework. Therefore, differences lie between unpaid housework and paid housework.

Housework division and cost-benefit comparison

For paid housework providers, the marginal utility of income is greater than that of leisure. Working as a labor to do housework, income increases, while leisure time and other activities time decrease. For the employer of paid housework, income decreases, leisure time and other activities time increase, but the marginal utility of income from doing his own housework should be less than that of leisure.

The internal arrangement of unpaid housework is often the result of the division of labor in society. It's important to determine the family's economic lifeline first, in order to determine who is better suited for housework and who is better for working outside? Or would it be better for the family to outsource the housework? In any case, there is an issue of cost-benefit comparison. When the income from going out to work is much higher than the expenditure of outsourcing the housework, a family will make the decision of outsourcing the housework. For example, the current development of China's express delivery industry is the result of the deep division of labor in society, which replaces part of the original housework.

When making decisions, people often make inner cost-benefit comparisons: to do something more worthwhile, labor productivity may increase, or the quality of life may improve, and welfare may increase. Therefore, in fact, unpaid housework and paid housework involve the transition from self-service to outsourcing, from a non-market approach to a market approach, which is a process of professional development and a change of the economic system. The increase in labor productivity will change the quantity of output. How to measure the increase in output brought about by specialization? Can we say that the change in institutional arrangements has nothing to do with production? This involves the principles that economic measurement should follow (Qiu, 2019a).

Unpaid Housework from the relationship between related activities

From the perspective of the relationship between housework, employment opportunities and leisure, different choices are made in a certain economic context. If family members' employment income is higher than the expenditure on housework outsourcing, and the employment opportunities are adequate, and other conditions are the same, people usually choose employment. If you work for \$20 a day, hire a labor for \$2. Of course, you choose to work and then hire a labor, so that the net benefit will increase.

But keep in mind that this happens when there are adequate employment opportunities while other conditions remain the same. When the job supply is insufficient and the possibility of unemployment is high, people should not take time off for doing housework or leisure, otherwise could be unemployed. Therefore, individuals and families should also have this counter-cyclical adaptive adjustment: working overtime, giving up leisure, and letting go of the housework that does not need to be done daily. Although housework and leisure time are reduced in a short time, the overall benefit of the family is positive in the long run.

According to the SSF Report, some people argue that the increase in leisure time should be recorded as a positive value. The author has doubts about this claim. Does that mean that the decrease in leisure time should be recorded as a negative absolute value? Just looking at one side is not enough.

In fact, the total benefit depends on the length of the observation period and the economic background. The observation angle is fixed and short-term to make an absolute judgment, without considering the impact of the economic cycle on the analysis object. It cannot be assumed that jobs are readily available, and the possibility of unemployment should also be considered. The relationship between different activities should be placed in the specific context of the economic cycle.

“Time Use Accounts” basis for unpaid housework estimation

To measure the output of Unpaid Housework, which is based on the measurement of time use, one needs to know the length of the work. The problem is that the basis of the measurement is not solid.

Time Use Accounts divides people’s time into five categories: paid work, unpaid housework, leisure (including sports, religious and spiritual activities and other leisure activities), personal care (mainly sleeping, eating and drinking) and learning. Then conduct a questionnaire survey and ask the respondents to fill in the form to find out how long each of these five activities take up in 24 hours? With the basic information, we can reprocess and summarize Time Use Accounts of different countries and make spatial comparisons.

The compilation of Time Use Accounts needs to implement the Principle of Mutually Exclusive & Collectively Exhaustive. In fact, it is also the Principle of No Omission and No Repetition emphasized in the original socio-economic statistics. After classification, each component item cannot be repeated or omitted. However, it is difficult to implement measurement in practice. Because the boundaries of these five kinds of activities are unclear and mostly “either one or the other”. Many people can do one job but belonging to more than one activity. For example, paid to play with a kid, a babysitter also engages in the activity of leisure besides paid work. The same goes for unpaid housework. Shopping in the store, of course, is doing housework, but it may also be leisure, or even learning—preparing for future employment. People cannot exactly divide their 24 hours into these five kinds of activities. As a result, the survey data and analysis results of Time Use Accounts are problematic, which will cause uncertainties, and those uncertainties will also be transmitted to the estimation of the output value of unpaid housework.

Take the case of commuting to work. Commuting should be an output of employees--they use their time (labor input) and money (commuting ticket) for this purpose. With the exception of the consumer’s purchase of a ticket for a commuter train, which counts as final consumption, this could be remedied by allowing for the household production of transportation services, which would be considered as an unpaid intermediate consumption of employees’ personal contributions to the output of the company. This interpretation is of course reasonable, but it can also have other economic interpretations, for its nature is very complicated. Many people would rather live farther away and take a long commute. For them commuting is uncomfortable for a shorter time, compared to living in a busy downtown closer to work, which is uncomfortable for a longer time. The housing choice as a welfare component may need to be at the cost of commuting, which is due to personal welfare preferences.

If unpaid housework is included, where is the boundary of output value calculation?

If unpaid housework is also considered productive labor, then what other activities are not considered productive labor?

Since the emergence of national income statistics, there has been a controversy over the boundaries of productive labor, which economic activities are productive, and which are non-productive? As we all know, in the history of economics, there are physiocrats, mercantilist, and later

narrow view of production. The narrow view of production emphasizes the material form basis of production. It was not originated by Karl Marx but inherited from Adam Smith. Therefore, the narrow view of material product system (MPS) comes from Western Economics. Later, Western countries expanded from a narrow view of production to a broad view of production. One of the unresolved disputes at that time was the production nature of unpaid housework.

Where does the boundary of productive labor end? If unpaid housework is included, are there only personal care activities that are not counted as production? Personal care is defined by the third-party criterion, and such activities must be carried out in person by each individual, not by a third party. The person concerned is hungry, thirsty and sleepy. If someone else eats, drinks and sleeps, instead of the person concerned, it will not only have no positive benefits, but may also have negative effects—it will cause more bad feelings, so these three activities are personal care activities.

It seems that personal care activities are impossible to be counted as production, but its boundary is blurred. For example, from the perspective of eating, improvement of living standards nowadays mainly depends not on what to eat, but on who to eat with. The bureau of the meal is very important. Whether it is business, government or academic exchanges, benefits can be generated from the meal. It is impossible to assert its non-production attributes. Is it possible to stipulate a two-hour meal, ten minutes is personal care, and the rest goes to production? Assume another extreme example. A prostitute accompanies the client for eight hours, and how many hours is personal care? How many hours are productive?

It shows that this issue, SNA is also facing the same dilemma as the MPS (the Material Product Balance Sheet System) of that year. The difference is only the boundary of production activities. Some people believe that SNA is a scientific accounting system, while the MPS is not scientific. In fact, this is a considerable misunderstanding. As far as the balance of the accounting system is concerned, the MPS is as scientific as SNA, and both conform to the tripartite equivalence principle of accounting, and self-balance relationship is established. The MPS has significant difference that only physical goods are counted while services are excluded, but in other respects it is an accounting framework similar to that of GDP (Coyle, 2014). Each of the two systems has value. They are intimately related because both focus primary attention on the production account, which is gross in one case (MPS) and net (i.e., adjusted for intermediate consumption) in the other (SNA) (Ward, 2004).

It is only for the modern service economy that has been fully developed, the production definition of MPS is too narrow to meet the requirements of modern economic macro-management, so it is necessary to switch to SNA. It should be noted that MPS is also an international accounting standard officially announced by the United Nations Statistical Commission. Especially in economic management of economically underdeveloped countries, MPS made its due professional contribution. Therefore, on this issue, historical nihilism should not be adopted.

Criticizing GDP without unpaid housework: the joke itself may be a joke

A university professor once criticized GDP grandiosely and told a quite classic story: A joke captures flaws with GDP measures by comparing the GDP effects produced by two individuals' activities. One is a happy family, where the male host goes home after work to his family. They get pleasure from cooking their gourmet meal together, using ingredients grown in their garden, followed by a quiet evening reading together. The net contribution to GDP is the value of the few ingredients for dinner that they purchased and the cost of buying books. By contrast, a bachelor eats junk food at a fast food restaurant, then goes to a bar to get drunk, and then visits a prostitute, and

then wrecks his car on the way home, then takes a taxi home. The latter contributes a lot to GDP—fast food expenses, alcohol money, prostitution, taxi and car repair expenses. This story uses moral standards to kidnap GDP and vividly illustrates the absurdity of GDP.

There is also the famous joke of “nanny and professor” from Paul Samuelson: what happens to GDP when a professor marries his servant? The nanny works for the professor, and the salary is included in the GDP. In the first year, they get married, and the nanny becomes a housewife. The professor no longer pays for housework, and the GDP decreases.

These two examples both comment on macroscopic things from a micro perspective but forget the trap of fallacy of composition emphasized by Professor Samuelson. Thrift is a virtue for individuals, but it can be a disaster the whole society. It is correct at the micro level, but not always at the macro level. On the contrary, what is right at the macro level may be quite wrong at the micro level. The inconsistency between judgment at micro level and macro level is not only possible, but quite a lot. Can the macro measure be evaluated only from the micro perspective? From the previous two examples, we should also see that bachelors provide opportunities for the realization of social productivity, while the frugality of warm men may objectively result in the idling of social productivity. Therefore, the judgment should also take the economic cycle background into account.

Some people say that disasters facilitate GDP, while others use it to criticize GDP. In fact, regardless of whether the benefits are positive or negative, the economic background should be considered. If the economic capacity is excessive, disasters may indeed trigger a turnaround. It turns out that the idling of economic capacity is a waste of social resources. After the disaster, the remaining economic capacity has a chance to be realized, and the value added may increase instead. But if the original production capacity is insufficient, it will be difficult for GDP to increase in the short term. Statically, the original resources have been transferred for use, and the economic scale will not increase. Therefore, it is necessary to consider the economic background.

People are often perceptual. It is a common way to understand macroscopic things from microscopic examples, while rationality is often achieved indirectly. Therefore, we need to consciously complete the transition of understanding from the micro to the macro. Just adhering to the atomic rationality, without taking the changes in the structure of things into account, it is enlarged to the macro level, which is derived from the micro reduction method from the methodological point of view. Following this way of thinking, what is wrong at the micro level is bound to be wrong at the macro level, but the influence of structural transformation in the process is not considered. From a more professional perspective, the two microscopic examples of ridiculing GDP have fallen into the trap of fallacy of composition, and in the macro sense, the jokes themselves have become jokes.

Are the two main methods of estimating output value of unpaid housework reliable?

Many people advocate the calculation of output value for unpaid housework. If we accept this idea, the next issue is how to calculate it. The focus is on measurement.

At present, there are mainly two methods for trial calculation. One is the market price reference method, which estimates the value of unpaid housework by referring to the market price of similar labor, multiplying average hourly wage rate of the market babysitter by the total unpaid housework time. The calculation is very simple. It's only a proportional calculation. But what assumptions are included? Are these assumptions valid? If the assumptions implied in the measurement deviate from reality seriously, what is the impact on the estimation results?

In the estimation of the market price reference method, it is necessary to assume that the market is vacant, and that housewives can find a nanny job if they intend to. Again, the salary level of the job you find is the same as that of the nanny who is already on duty. Of course, the salary level of nannies in different positions is different. Usually, it is calculated by the average, but can it be estimated in this way? Other factors need to be considered: after more housewives enter the nanny market, the level of labor supply increases. Will this lower the salary level of nanny?

Some may say that the supply of labor has increased, so has the demand. Because the housewife goes out to work, her housework also needs external employment. The question is whether this change in supply and demand can be balanced. In addition, what is the dynamic impact of changes in supply and demand? Should the degree of specialization of housework be improved, but with the increase of efficiency, will the wage rate change? Therefore, the above assumptions are not necessarily so reliable.

Due to the various uncertainties in estimation method of market price reference, it has been proposed to estimate according to the opportunity cost method. There is another reason why the algorithm needs to be changed: with the progress of society, many families do not have housewives alone to do housework but different family members in turn. The concept of housewives has fallen behind in developed countries, which requires estimating output value of unpaid housework of each family member separately.

The so-called opportunity cost method regards the wage rate of unpaid housekeeper's original job as the opportunity cost of abandoning the original job to engage in unpaid housework, multiplying the loss of income per unit time by the unpaid housekeeping time, thus obtaining the estimated output value. For example, a famous professor does not know how to cook but makes a meal for his family. No matter how bad it tastes, the output value of the professor must be calculated according to the professor's salary level.

But we know that the efficiency of professors engaged in knowledge production is far from that of housework. To estimate the value of unpaid housework output according to the opportunity cost method, it is necessary to assume that the efficiency of people engaging in different positions is the same, and the income is the same, so that the opportunity cost or loss of income may be the same. Obviously, this is quite different from economic reality. And estimates can easily slip into absurdity.

However, it may be reasonable to estimate according to the professor's salary level. Although the meal tastes bad, after all, it is made by a famous professor. The family members may feel that he is so busy, but still makes the dinner with love. It is a highly scarce welfare product, unpalatable and happy. If the output value measurement should focus on well-being, which is very effective to improve well-being level. Can we understand the estimation of opportunity cost method in this way? Maybe it can, perhaps it is a reasonable perspective.

At present, these two methods are mainly used to estimate unpaid housework, and the rationality of other methods is even worse. However, the author believes that neither of these two methods can meet the standard of the conventional statistics. Even if housewives' unpaid housework should be calculated, from the point of view of feasibility, it is not necessarily possible to incorporate it into conventional economic statistics at this stage.

From the perspective of economic operation, the market has a strong ability to choose and absorb, and it is bound to incorporate those economic activities and resources that are important to society into its own system. If it has not been included in the market, it means that this factor is not so important to society. If we believe in the market, the main body of economic activity is the part that has already entered the market.

After this part of measurement is accurate, the mission of national accounting is basically achieved. Banks often assess the capital adequacy ratio and the core capital adequacy ratio. The calculation of these two related indicators is simple, but the idea is very important. That is, where should the focus of measurement and accounting be placed? Imagine two measurement choices: pursuing comprehensiveness and far from accuracy, grasping focus and being relatively accurate. If we need to decide, how do we measure it?

2.1.4. Identify doubts in “Regretful Spending”

In 1973, Professor J. Tobin and Professor W. Nordhaus published a very famous paper, emphasizing that the measurement of production results should focus on welfare. For this reason, they put forward the concept of regretful expenditures (the SSF Report calls it defensive expenditures), which many people in the industry regard this as a precedent for GDP improvement. The authors believe that there has been a contradiction in the macroeconomic measurement. Tobin and Nordhaus uncovered the lid. Many measurements and accounting items have unclear boundaries. There is no Berlin Wall between necessary and non-necessary expenditures.

Are defense expenditures and prison expenditures “defensive”?

At the same time, difficulties abound when it comes to identifying which expenditures are defensive and which are not. The park construction expenditure mentioned in the SSF report, does it belong to regrettable expenditure (to alleviate the discomfort of urban life) or should it be counted as final consumption (public entertainment services)? However, the SSF Report clearly regards defense expenditure as defensive expenditure (Clerc et.al, 2011). The author argues that there is no clear boundary between related items, considering defense expenditure as a defensive expenditure raises joint doubts. An important consideration for grouping in economic statistics is where the boundaries end. The key is how to deal with its extension effect or ripple effect. If defense expenditure is regarded as defensive expenditure, then many similar expenses will be defensive expenditures.

Can defense expenditure increase people’s well-being? The key is to judge the selected perspective, and we can find obvious reasons for positive affirmation. Because the national defense is strong and people live and work in peace and contentment, and well-being itself has increased. Why do Americans always say God bless America? This sentence is essentially based on a strong military force. Is it not productive that the Chinese Navy is escorting in the Gulf of Aden and corporations can safely transport goods? According to “Understanding the National Account”: they clearly contribute to production, since there would be much less output if the government failed to defend the country against a foreign invasion, to maintain law and order and to keep the road system in good condition. Indeed, the absence of such services can lead to a catastrophic decline in output, as the experience of numerous developing countries prove this (Lequiller and Blades, 2014).

In addition, positive benefits of prisons: more families can live in a relatively safe environment with fewer criminals, a better consumption mentality, and a direct improvement in living standards, which is an improvement in the community environment. In addition, households have anti-theft windows installed in their homes, so that they can sleep safely. Is it an intermediate input or final consumption? What’s the difference between anti-theft windows and sleeping pills? If installation of anti-theft windows is defensive expenditure, then is buying sleeping pills? If buying sleeping pills is defensive expenditure, then buying melatonin? Where do we have to “regret”?

The influence of direct and indirect effects on defensiveness

Sometimes it is not necessary to directly reduce the loss, just to reduce the risk of loss, it's welfare itself. The more we consider the shift from the measurement of production results to the measurement of welfare, the more we need to consider the indirect effects between things. Some people ridicule the term "negative growth". Why do economists talk in a roundabout way? The author's explanation of economic measurement is that if we want to integrate the benefits and losses and include them in a measurement boundary, we should ignore the basic point of zero, which makes sense. It is not necessary to make an absolute distinction between positive and negative utility. If potential losses can be reduced, it means potential benefits.

Moreover, if defensive expenditure should be determined, the estimator needs to take a stand and make moral judgments. The concept of defensive expenditure is inconsistent with the Measuring Moral Non-involvement Principle and the Neutral Requirement of Economic Measurement of economic statistics (Qiu, 2019a).

How to calculate health expense? Is it counted as human capital investment or final consumption? Many people prefer the former. The famous fable, Fish Man and Rich Man's Seaside Dialogue, shows that the two people have different preferences, different perspectives and different qualities of behavior.

Is physical health an end or a means? In fact, there is a duality. If we focus on the long-term, then personal consumption can also be an investment in human capital. In Japan, eating beef is an important measure to improve physically. When Japan's economy was underdeveloped, cattle were only used for farming. Killing cattle was destructive to production and cattle were not allowed to eat. Later, only by improving physically can they advocate eating beef.

So, is beef consumption as final consumption or human capital investment? The new measure is recommended for consideration. According to the original national accounting, eating beef is final consumption. If we pay attention to human capital investment, eating beef has the nature of intermediate consumption. The focus of the two considerations is different. If the estimation of human capital is included in the SNA, how much will final consumption remain in the original national accounts? How to balance the issues solved by the new method and the problems it brings? Human capital accounting has not yet been included in the SNA. The reason is that the two measurement perspectives cannot be coordinated.

Take another look at the issue of the extra consumption of food. To do heavy physical work, you must eat more. Suppose you originally eat two steamed buns, and now eat four. If the excess food is included in final consumption, GDP will rise, but the person concerned does not feel that he has consumed more, and welfare has not increased. Whether is defensive expenditure to eat these two steamed breads so much? In addition, it is impossible to determine whether the extra two steamed buns can compensate for the consumption of physical work. The problem is to examine the cost and benefit over a long period of time.

The Extra consumption of food is only a micro-introduction, and the focus of economic measurement is on the same macro-level. In contrast, some people believe that putting more criminals in prison does not make society better, but this judgment implies the hypothesis that there is no increase in violence. Otherwise, in real environment, the increase of prisons objectively actually compensates for the losses caused by the deterioration of the social environment. Even if the loss is potential, it may not be real. If you sleep well, you don't need to take sleeping pills. Can sleeping

pills therefore be regarded as defensive expenditure? In contrast, there are similarities in measurement between increasing prison consumption and eating more bread. There are many expenditure items like this kind of ambiguity. How to define the boundary between defensive and non-defensive expenditure?

Welfare can be divided into material welfare and non-material welfare. If we only consider material well-beings, the concept of defensive expenditure may be valid, but if we need to consider non-material well-beings, defensive expenditure may not be defensive. Looking at the case of the increase in prisons, the basic point of social well-being comparison is that a society with increasing violence and a society without increasing violence. The object of social well-being evaluation is a society with defensive expenditure and a society without defensive expenditure. The social environment changes, and the relationship between supply and demand changes, hence the basis of comparison changes, the object of evaluation changes, and the measurement changes accordingly.

How to improve the measurement to avoid the trap of defensive expenditure?

In response to the so-called defensive expenditure, the SSF Report puts forward three recommendations: focus on household consumption rather than total final consumption; widen the scope of assets; widen the scope of household production. The biggest obstacle to these approaches lies in the feasibility of implementation. How exactly should the scope of defensive expenditures be determined? How should new assets and in-kind flows be valued? And, of course, widening the scope of assets and production measures brings with it more imputations. If these issues are not implemented, the recommendations made in the SSF Report are castles in the air.

The four measurement dilemmas described in this article are not all the hidden dangers of GDP statistics. There are many problems worthy of in-depth discussion in GDP statistics, such as the financial intermediation services indirectly measured (FISIM) method of banking services measurement. It has been included in the SNA, but after the 2008 financial crisis, some economists have strongly questioned this innovative method. In addition, the GDP statistics should also be summarized and summarized from the perspective of general methodology and the problems and countermeasures.

What does it mean to focus on household consumption? Economic measurement is not only macroscopic, but also microscopic. But what are the potential measurement problems of household as the basic unit of measurement? Households of different sizes are also different as consumer units. Moreover, there will always be final consumption in society. Some public information that cannot be decomposed into households or individuals, such as lighthouses, streetlights, and public information on official websites. Can these public goods be recorded as defensive expenditure? Also need to consider: is it to supplement the household perspective, or to replace macro perspective?

The problem is that the logical connection of concepts does not imply the feasibility of reality measurement. The links between the indicators, including the original indicators and alternative indicators, need to be measured logically. If the improvement method hides more problems than it reveals, or brings more problems than it solves, the improvement is worthy of in-deep thinking.

Further exploration is needed: are the questions raised above already included in the SSF Report? Is their consideration of this special issue so thorough? If the author's doubts can be established, what will be next improvement recommendations?

The four measurement dilemmas described in this paper are not all hidden risks of GDP statistics. There are many problems worthy of further discussion in GDP statistics, such as the financial intermediation services indirectly measured (FISIM) method of banking services

measurement, which has been incorporated into the SNA, but after the 2008 financial crisis, some economists have strongly questioned this innovative method. In addition, the GDP statistics should be summarized and induced from the perspective of general methodology.

3. Exploring the feasibility and necessity of well-being measurement

The measurement of well-being is a long tradition in economics and economic statistics. Jeremy Bentham, a British philosopher, creatively compiled the Happiness Index as early as the 17th century. The German “School of Staatenkunde” and then later “School of Social Statistics” also paid special attention to social well-being. In the debate on GNP or NI, Simon Kuznets, an American economist and Nobel Laureate in Economics, emphasized that economic measurement lies in social welfare (Coyle, 2014). After the establishment of the SNA paradigm, Richard Stone designed the system of social demographic statistics (SSDS). Since the beginning of the social indicator movement in the United States in the 1960s, well-being measurement has become a hot topic of social concern. Various measures have emerged, and various measurement methods have become popular. This tradition is so powerful that people tend to regard measurement of well-being as justified.

However, is well-being measurable? Are all popular methods of measuring well-being practicable? Is the measurement logic on which it relies really linked? Even, why is it necessary to measure well-being? At least, what kind of economic well-being measurement content is necessary? Under what cost constraints is it necessary? Whether it is the SSF Report or related literature, there is little systematic discussion on these two basic issues, and it seems that feasibility and necessity of well-being measurement is not a problem. The author disagrees.

This part is the author’s questioning and criticism of well-being measurement based on content of the SSF Report. The first issue discusses the diversity of economic well-being and its possible obstacles to the measurement, as the object of the measurement method. The second issue analyses three conceptual approaches of measuring economic well-being: subjective well-being (SWB), the notion of capabilities and the fair allocation approach, focusing on the methodological enlightenment of these concepts. The third issue reveals the feasibility of main methods of well-being measurement, including questionnaire method, willingness-to-pay method, representativeness of samples and composite index method. The fourth issue points out the four deficiencies of the SSF Report from the perspective of the text: First, the insufficient logical links between concepts and specific methods; Second, the systematic exposition and comparison of various measurement methods; Third, the feasibility analysis of the methods; Fourth, the status of monetary aggregate method in well-being measurement. In addition, the feasibility of biotechnology in well-being measurement is also discussed. The fifth part discusses the feasibility of economic well-being measurement from the perspective of public goods boundary, policy significance and information tax.

3.1. Multi-dimensional well-being as a measuring object

3.1.1. The definition boundary and pattern of measuring object

The extension of well-being is not easy to grasp. The SSF Report uses the restriction of economic well-being and equates it with the quality of life. As for the difference between well-being and economic well-being, the SSF Report is unclear. Nor does the SSF Report use the popular term happiness

measurement. These prudent practices warn us that once the field measurement is involved, we need to pay attention to the appropriate grasp of the connotation and extension of economic well-being.

Economic well-being measurement is different from GDP statistics. Instead of focusing on the level of production development or economic performance, it shifts economic measurement from the production process to its purpose. However, there are differences and connections between well-being measurement and GDP statistics.

It should be noted that the SSF Report highlights the distinction between current and future well-being, that is, the quality-of-life measurement should be distinguished from the sustainable development measurement. The author believes that this distinction is quite necessary from the perspective of measurement feasibility and data quality. Economic measurement will always face various constraints. Separating the more assured measurement from the less assured measurement is a manifestation of being responsible for data users, and it is also a kind of awe in economic measurement.

The author puts forward a hierarchical distinction: the three major contents of the SSF Report constitute the three levels of economic measurement (Qiu, 2019b). Our understanding of economic well-being measurement should be understood in the overall pattern of economic measurement.

3.1.2. The difference between production measurement and well-being measurement

There are differences between production measurement and well-being measurement. Production is for well-being, but it is only one of the sources of well-being. From the perspective of measurement, well-being sources can be divided into four kinds: First, the unit's production in the current period. Second, the transfer of production results between different time periods, such as the positive externalities of public services exerting a lagging influence in each period, which should even include the missing part of the results in each period. Third, the spatial transfer of production results and the positive externalities of one country to another, of course, there are also negative spatial externalities of well-being transfer. Fourth, the effect of natural endowments such as natural resources and climate on economic well-being.

The first three kinds of economic well-being come from production, but their time and space range are inconsistent, leading to the difference between well-being measurement and production measurement in a specific time and space. The fourth type of well-being does not come from production activities, such as a city being warm in winter and cool in summer, which can save resources of air-conditioning equipment and its operation, which can be used to meet other well-being needs.

The influencing factors of well-being demand can be subdivided: First, the consumer psychology of the demand side. Individual consumption habits are different. There are two basic consumption habits, pessimism and optimism. One is to pursue the best spot and enjoy the best every time; the other is to pursue the best expectation and enjoy waiting each time. For example, there are two ways to eat grapes, from good to bad, or vice versa. Second, the original well-being function of the demand side. The distribution of objective well-being demand in different periods can produce a certain offset effect. If the foundation is good, short-term well-being is worse and can be tolerated. If the supply is reduced, well-being level will not be affected. But if the original well-being function is not good, the economic situation becomes worse, or the well-being level is unable to improve for a long time, it will be unbearable. Third, the sense of satisfaction and dissatisfaction generated by different spatial comparisons on demand side. Fourth, the Influence of culture, religion and history, etc. For example, the Puritans emphasize frugality, and some extremely frugal sects did not even use modern electrical appliances. Only when they live a primitive life can they be truly happy.

Factors of supply and demand can and should be subdivided, and the combination of the two factors will produce multiple rounds of cross-effects. All these make well-being measurement and production measurement quite different.

3.1.3. Multi-dimensional well-being and limited measurement

Well-being is multi-dimensional, but measurement resources are limited. Only part of the content can be selected to reflect it. No matter which part is selected, it will be a kind of paranoia. The paranoid well-being measurement will become a kind of information temptation, which will lead to people's flattering psychology and conformity psychology. Temptation objectively interferes with people's happiness orientation and destroys multi-dimensional well-being pattern, which is a kind of cultural autocracy. Some people in Europe and America admire Bhutan's gross national happiness (GNH). In fact, Bhutan's happiness is of material hardship, based on the religion of pursuing afterlife. How many Europeans and Americans can live in it for a long time? Can the people in developing countries be comforted with hardships? What is the positive significance of such well-being measurement? Today, some Bhutanese themselves are not willing to completely replace GDP with GNH.

Cross-cultural comparisons of economic well-being may not be entirely feasible. Different cultures have different preferences for quietness and liveliness. Developed countries pay attention to quietness and create a low-noise environment by means of public morality, law and technology, for example, Japanese are not allowed to talk on mobile phones on buses. Different cultures have different dietary preferences and taboos. Traditional Chinese medicine benefits people in Eastern culture, while many people in the West consider cupping, scraping and bloodletting as witchcraft. The relationship between parents and children is also quite different in different countries, which is reflected in whether to scold, how much to support, and the cost of marriage, and so on. Europeans and Americans believe that it is poor leisure that individuals, families or children cannot afford a week of holidays away from home at least once a year. But most people in poor countries do not have this luxury. Holidays are often time for non-routine housework. If they can do nothing and really rest their bodies, they will be very satisfied. There is no poor leisure at all. It shows that different countries have different demand and opinions on the quality of life at different levels of economic development. The well-being increase in the East may become the well-being decrease in the West, which cannot be measured by a unified standard. Thus, what is the additivity of different types of well-being?

In terms of the international comparison of well-being measurement, do countries at different stages of economic development have the same degree of demand for such public goods? Should a country's economic measurement match its economic level, or should it follow international standards? This is a practical question. For Europeans and Americans, they are almost equivalent, but can that reflect the needs and capabilities of poor countries? Should standard economic well-being measurement requirements be put forward? At what level of development can the requirements of deepening well-being measurement be proposed? But if the requirements for well-being information are different, for example, the classification of well-being information is different, how to construct aggregative indicators for in-depth international comparison?

3.1.4. Is leisure measurable?

Leisure is an important component of the quality of life. The Organization for Economic Cooperation and Development's "Understanding National Accounts" defines leisure as "time spent on activities that are unrelated to production and personal care. (such as eating, drinking and sleeping)" (Lequiller and Blades, 2014). In fact, this statement is not correct. It is difficult to separate leisure from production and personal care. It is also entangled with unpaid housework. Even if it is not directly related, it can also be indirectly related. It is this characteristic of leisure that causes the measurement dilemma. Whether it is the quantity of leisure time or its quality, it is difficult to give an exact measurement result. To measure, there must be corresponding standards, while leisure varies from person to person, event, time and space. People often fail to formulate a unified measurement standard.

Leisure is related to people's abilities. People with strong abilities may spend part of their working time at leisure. People with weak abilities work overtime after work. Leisure, ability and work management methods can interact. Work can be divided into two types: piecework and timework. If timework management is adopted, people with strong ability should consciously reduce labor intensity to maintain a balanced work rhythm. Just imagine how low labor intensity can be regarded as leisure.

Leisure is related to people's mental state. People are different. Some people always keep oneself busy, uncomfortable without work, and often worry about having too much leisure. Some people are very afraid of things. They are still worried about how they are doing their work during non-working hours. They can't be in a state of leisure at all.

Leisure is related to people's attitude towards life. If work and personal hobbies can be combined, what is the status then? It is both work and leisure, and it's impossible to decide. Many market jobs can overlap with leisure, such as painters, poets, researchers, etc. Inspiration often comes from leisure process, from natural phenomena, from social interaction process such as eating and drinking tea, etc. Intellectual workers may not work overtime, just because they never leave work mentally. The "double helix model" of human genes was born in the Eagle Bar, Cambridge, England. The "Laffer Curve" of economics was also drawn on a napkin. Many people envy the profession of university teachers, who have two vacations a year. In fact, the characteristics of teachers' posts lie in the flexibility of working hours, which makes it difficult to distinguish between leisure and work. Since leisure is easily confused with other activities, we should not only look at the superficial phenomena, but also make a thorough investigation and pay attention to the distinction.

Leisure may also be a necessary intermediate consumption in the production process. For example, high-tech companies in Silicon Valley in the United States have set up leisure areas in the buildings, and it is mandatory for employees to go to leisure during work hours. They can relax in various ways, otherwise, their work efficiency will be lower after mental fatigue. The more high-tech work, the more mentally intensive work, the more adjustment is needed. Highly stressful mental work requires employees to take regular vacations every year, and after work on weekdays, they need to engage in arts and sports to relax their nerves. In this way, the so-called leisure activities may also be intermediate consumption paid for the production process, which is individual's contribution to the production of the enterprise, which can be so qualitatively so.

Leisure is likely to be confused with unpaid housework. Whether some activities are leisure or unpaid housework depends on personal likes and dislikes. Some people enjoy cooking while others are willing to go shopping. Some enjoy surfing the Internet and others enjoy driving for pleasure. For

those who are full of enthusiasm in family life, unpaid housework is just leisure. The difficulty in well-being measurement lies in: how to allocate time to the two?

Leisure may also be confused with “personal care” (mainly eating, drinking and sleeping). If a meal lasts two hours, how much time it considered personal care? How long is it leisure? Going to bed and going to sleep are usually two different things. Is lying in bed considered personal care during this period? Or leisure? Or work? Or study? It depends on what the brain is doing at that time, whether it is counting sheep, “Xintianyou”, which is a Shaanxi local melody in China, thinking about the work plan, or reciting Tang poetry. During this period, not only the state is difficult to classify, but also the state is often changing and cannot be defined.

The author summarizes the doubts about measuring leisure into three points: First, Leisure and other human activities have too much “one and the other” to satisfy the principle of “neither repetition nor omission” in statistical grouping. It is impossible to determine how much time an individual spends on leisure every day. It cannot be measured clearly (Qiu, 2019a). Second, the quality of leisure is also difficult to measure. For different people, leisure cannot be evaluated uniformly, and its “equivalent income” can hardly be calculated according to one standard. Third, surely certain data can always be calculated for leisure. The problem is that this calculation contains too many assumptions and the result is plausible. It does not have the socio-economic connotation that it should have. It cannot really reduce the uncertainty in social cognition and may even mislead people’s social cognition.

In short, leisure is difficult to measure, and as an indispensable part of human life, the immeasurability of leisure will seriously affect the exact measurement of the quality of life.

3.1.5. The inequality measure and the impact of the increase in its maximum value

The existence of inequality itself is the decline in the overall quality of life in society. Inequality is more multi-dimensional in economic well-being. Different types of inequality have their own meanings and cannot be replaced by a measure of inequality (such as income inequality indicator). Moreover, inequalities may reinforce each other and produce superimposing effects, which requires research on the impact of policy measures on their effects. The SSF Report emphasizes the cumulative effects of multiple disadvantages, for example, the loss of quality of life due to poverty and disease is far greater than the sum of the losses caused by the two (Stiglitz et al., 2010).

The author believes that, correspondingly, we should also pay attention to the cumulative effect of multiple advantages. For example, it is easier to start at the workplace for individuals to be both good-looking and capable than only having either one of the two advantages. The combination of the two situations is known as Matthew Effect: the more you have, the less you are deprived. Therefore, both aspects should be grasped at the same time. When we research inequality, we should not only look at the accumulation of disadvantages, but also add the accumulation of advantages.

Special attention should be paid to the phenomenon of “red shift” in nature, and the distance between planets is expanding. The author believes that there are similar phenomena in the economic circle, and there is also a trend of maximum increase in economic scale and related indicators. The unilateral expansion of the extreme value leads to the enlargement of the distance between the evaluated objects and of the pattern of things being evaluated.

In the 1980s, when China replayed the old film “The Million Pound Note”, which shocked people in China. How could a foreign millionaire have millions of pounds? That is wealth that Chinese people could not even imagine in the past. With the development of economy, wealthy

people become increasingly richer. Some depend on family inheritance and more on personal struggle, such as Silicon Valley Technology elites and Wall Street Financial elites, whose wealth is unmatched by the industrial giants back then.

To compare wealth of different ages, of course, the factor of inflation should be considered. A Chinese multimillionaire of today is not equal to a British millionaire of the past. The entire economic history is a history of inflation, and nominal wealth is always increasing. However, eliminating the factor of price, it still shows a change of more than orders of magnitude. It should be noted that due to the expansion of the economic order of magnitude, the degree of objective inequality will certainly increase.

Therefore, in the time series analysis of inequality, there are incomparable factors for different levels of inequality. An increase in the value of the same inequality index may not necessarily be the increase in inequality, which will have the effect of a maximum increase. When calculating the relative number, the absolute value of the 1% difference is different, and the actual economic meaning of it needs to be considered. Just as the economic cycle analysis needs to exclude the impact of long-term trends, the spatial comparison needs to exclude the impact of different stages of development, and the inequality analysis also needs to exclude the impact of order of magnitude variation. How much is the increase in inequality caused by the increase in maximum value? After deducting this influence, the inequality level that is comparable in time and space is the level of inequality that is brought about by structural changes. The SSF Report pays attention to the issue of inequality (Stiglitz et al., 2010), but fails to recognize it. The author believes that the factor of maximum expansion must be fully considered in analysis of inequality.

3.2. The three conceptual methods for measuring economic well-being

Measurement of economic well-being should be based on philosophical ideas, because philosophical thinking has a long tradition of focusing on what gives the quality to life, so which criteria should be used to evaluate the quality of life depends on the philosopher perspective that people adopt. The SSF Report summarizes the three main concepts of economic well-being measurement. The first one is based on the notion of subjective well-being, the second on the notion of capabilities, and the third on the economic notion drawn from well-being economics and from the theory of fair allocations.

3.2.1. Subjective well-being

In GDP statistics, monetary aggregates method mainly evaluates people's trading behavior, and it assumes that different trading choices show the preferences of economic entities. For a long time, economists assumed that it was enough to just look at people's choices to derive information about their well-being, and that these choices would conform to a standard set of assumptions. The SSF report pointed out that in recent years, however, research has focused on what people value and how they act in real life, and this has highlighted a large difference between the standard assumptions of economic theory and real-world phenomena. A significant part of this research has been undertaken by psychologists and economists based on subjective data on reported or experienced well-being.

An important function of subjective well-being measurement is to make up for the insufficiency of revealed preferences information. This measure divides the quality of life into three aspects: cognitive evaluation, positive feelings, and negative feelings. By using the questionnaire method, the respondents' answers to related questions are used to observe their preferences and self-identification.

One characteristic of the subjective measures of quality of life is that people's responses to their own situation do not have an obvious objective counterpart. We can compare the perceived and the actual inflation rate, for example, only respondents can provide information on their own subjective states and values, making the measurement reliability and comprehensibility compromised. In contrast, the perceived inflation rate and the actual inflation rate can correspond to each other, and the values of the indicators experienced by Individuals and calculated can be mutually confirmed.

Hedonic introspection refers to subjective measure of one's own happiness or utility by the individual. Studies have proved that this method has relatively stable validity and reliability (Wilkinson and Klaes, 2012). Reliance on the individual's own judgement is a traditional philosophical view. The so-called "whether shoes are comfortable is only known by one's feet" emphasizes this meaning, but there are still many problems in subjective well-being measurement and cannot completely replace other methods. Professor Nick Wilkinson points out that there is a fundamental distinction between individuals' continuous hedonic experience and meta-awareness or meta-consciousness. According to Bem's Self-Perception Theory, people often lack meta-awareness of their own internal states, resulting in the tendency to infer states, attitudes and preferences from their behavior. This process is subject to considerable error, in terms of misattribution (Wilkinson and Klaes, 2012). The SSF Report also points out that these memories and evaluations, however, can also lead to systematic errors (Stiglitz et al., 2010). This possibility demonstrates that well-being measurement guided by this philosophy may fail to guide well-being measures, at least partially. Therefore, it can be questioned that the mechanism foundation of subjective well-being concept is not very solid.

In behavioral economics, the concept of self-serving bias is put forward. A common example is so-called "better than average effect". People often think that they are better than the average, at least they should be. In this way, one's subjective feeling is often better than the actual situation, and when being negatively impacted by reality, the negative emotions caused by being far away from expectations will become more serious. These two possibilities may lead to the deviation of subjective well-being measurement from the actual situation. It should also be noted that the subjective well-being measurement emphasizes the distinction between positive and negative feelings. However, behavioral economics regards negative feelings as a guarantee mechanism and a double-edged sword. The impact of this on the subjective well-being measurement is that the first-order differentiation between positive and negative feelings is of little significance, or that every level of interaction between positive and negative feelings needs to be distinguished.

Subjective choice is based on memory and value judgment. But it may lead to bad choices, and some choices are made unconsciously, rather than by weighing the pros and cons of all available alternatives. Here we need to pay attention to the time limit issue of decision-making. Both selection and decision-making are subject to time constraints. Whether decision-maker is ready or not, some decisions must be made within a given time window. Decisions that seem to be irrational, if time-limit factors are added, are likely to be rational, even if they are decisions made subconsciously, even though people do not know they know it.

Do not misunderstand that there is no subjective measurement before. If one carefully examines the history of economic statistics, he will know that subjective measurement is a traditional tool. Many characteristics of economy and society are measured by people's answers to a set of standard questions. For example, Jeremy Bentham's calculation of Happiness Index in his early years requires the understanding people's subjective feelings. Looking at modern unemployment statistics, the questionnaire method has played an important role, mainly involving the three whethers: whether

one is working in a certain period, whether one is actively looking for a job, and whether one is in a state where can start working in the near future (Stiglitz et al., 2010). Of course, this is just a basic question. There is also the expansion question, such as the amount of time spent working in a month or so. Only when the accumulated working time is less than the specified number of hours in a certain period can it be considered as unemployed. The problem is that as we rely more on subjective measures, we should be more careful about the hidden obstacles.

Behavioral economics, cognitive science and social cognition have made great progress in recent years, and there are many theoretical disputes. Subjective well-being measurement should draw on the experience and achievements of these basic disciplines as an important research direction of economic measurement methodology. Since the basic disciplines are still developing, the measurement methods supported also need to accept new challenges from time to time.

3.2.2. How to sum up the objective features required by the capabilities approach and the fair allocations approach?

Both the capabilities approach and the fair allocations approach pay special attention to the objective conditions of people's lives and the opportunities available to them, which are the basis for calculating well-being indicators. Both the capabilities approach and the fair allocations approach belong to the multi-index comprehensive evaluation, that is, a series of component indicators are condensed from the selected well-being influencing factors, and then combined to obtain an overall evaluation.

So, what factors should be included in the list of objective characteristics? Is it to assess the changes in a country's domestic living conditions or to compare the living conditions of countries with different levels of development? The selection of influencing factors depends not only on the purpose of evaluation, but also on the value judgment of different economic entities.

Since the selection of factors affecting well-being also requires subjective value judgment, can the selected factor set reflect objective reality? According to the SSF Report, in practice, the practical experience of these deliberations has highlighted some themes shared across by many constituencies; a similar degree of consistency has also emerged when comparing the frameworks developed under various initiatives that focus on measuring broad concepts, such as well-being, human development, and societal progress (Stiglitz et al., 2009a). For example, it emphasizes the impact of social organization on people's lives.

However, the author believes that we need to think deeply about the source of this consistency. A basic fact that cannot be ignored is that the formulators of well-being measurement rules and the selection of measurement factors mostly are from developed countries. Does it include their subjective components? Only when measurers from various countries fully discuss, cross each other, with multiple rounds of feedback, and concentrate on subjective impressions of various well-being influencing factors, can they truly converge to the objectivity.

The SSF Report lists eight objective features shaping the quality of life: health, education, personal activities, political voice and governance, social connections, environmental conditions, personal insecurity, and economic insecurity (Stiglitz et al., 2010), and explains the main points of its measurement one by one. The author believes that there are still many problems to be further considered in the measurement of these factors.

The SSF Report admits that human health contains many different dimensions. Although several comprehensive indexes have been designed, attempting to measure human health, none of them are

universally recognized, which inevitably depend on the controversial ethical judgments and the weights given to different medical conditions. Whether traditional Chinese medicine or Western medicine, it is difficult to agree on this difference of judgment. The question is: can comprehensive evaluation be realized at this level? If the answer is no, what does it mean for higher-level synthesis?

The health measurement can also be expanded from different perspectives. For example, setting scenario A, the health quality of the bottom population is improved by 3%, the health quality of the top population is decreased by 1%, the health quality of the total population is increased by 1%. Can this result be accepted by the society? The reality of well-being measurement is that quality evaluation is often dominated by the top-level population. Then Setting scenario B, which comes first, length of life or quality of life? If the quality-of-life measurement is considered comprehensively, how should the standard be established? How to judge euthanasia and coordinate it with the existing measurement indicators?

Europeans and Americans usually value democratic politics. In fact, political voice and governance is only a moderate indicator, which implies a paradox: if the agent is reliable, no need for the public to participate too much; on the contrary, if the politics is inherently unhealthy, many people tend to stay away from politics. European and American countries believe their level of democratization is the highest, but people are sometimes reluctant to exercise their voting rights, showing that the highest level of democratization is not necessarily the best democracy. In addition, should we focus on political development at a comparatively similar economic level?

Rich country's elites believe that social connection covers a lot, so various indicators need to be designed for measurement. The fact that people of different cultures and strata have different needs and evaluations for solitude and social connections should be considered. Therefore, these indicators of social connections are not necessarily as large as possible. Does the larger value indicate harmonious social connections? Does more people living by Walden Lake indicate deteriorating social connections? When designing, calculating and interpreting the component indicators, attention should be paid to the differences in different cultural groups.

Measuring Job Insecurity requires assessing the security level of the individual's present job. What needs special consideration is: how to distinguish between upward instability and downward instability? Because the promotion of workplace also shows instability, but it is not that the work is insecure. Instability is not equivalent to insecurity.

People in different countries and at different stages of development have different well-being preferences and attach different degree of importance to different features. We should also consider: are these eight aspects comprehensive? What other aspects need attention? Each aspect of the features that need to be paid attention to when measuring well-being should be discussed by measurers from various countries, instead of directly being prescribed by European and American experts, for other countries to follow. In addition, we should not ignore the feedback mechanism in design optimization.

3.3. The main methods of measuring economic well-being

Based on different well-being measurement concepts, various well-being measurement methods have been developed accordingly. There are two basic methods used to measure well-being at home and abroad. One is the questionnaire method, which mainly measures subjective well-being; the other is composite index method, which is used in the capabilities approach and the fair allocations approach to synthesize the effects of various well-being factors.

It should be fully noted that these two basic measurement methods, including GDP correction method and biotechnology method, have various defects in measurement logic, which require further correction and vigilance of data users.

3.3.1. Questioning the questionnaire method

The questionnaire method is mainly used to measure subjective well-being, and it can also provide supplementary information for the capabilities approach and the fair allocations approach. For example, activities (such as commuting, working, or socializing) may be more important for the affects, while conditions (such as being married, or having a rewarding job) may be more important for the evaluation of life. In both cases, however, these measures provide information beyond that conveyed by the income. For example, in most developed countries, young children and the elderly report a higher evaluation of life than those of prime-age people, which is in sharp contrast with the income level of people of the corresponding age groups (Stiglitz et al., 2009b). It is not that the higher income, the happier and the better the evaluation of life. In addition, different people have different opinions on social affairs such as going to work, commuting, social interaction and marriage, and the messages they convey are also different.

There are also limitations and risks in using the questionnaire method to measure economic well-being, which are mainly manifested in the following three aspects:

The first is the requirement of the questionnaire method on the cognitive ability of the respondents. Not everyone is able to perceive the quality of life. Questionnaires can only be sent to adults with normal conscious awareness. But how to tell whether people with normal conscious awareness or not? Is there an absolute boundary between the two? The enormous pressure of modern society makes people mentally unwell to some degree. Rich or poor the country is, there are people with mental disorders. So, to what extent should the disorder be disqualified from answering the questionnaire? Who can absolutely prove to have the qualified mental system to answer the questionnaire? How to complete the confirmation and conversion of eligibility to be investigated between normal people and mental patients? The so-called autism patients, who are regarded by the society as abnormal just because they do not communicate with the outside world. But who can say for them that they are unhappy? Autism makes talented artists, showing that their spiritual life could be very rich. Perhaps it is a happy loneliness, a world where people are unable enter. So how can we jump to conclusions?

There are also minors and the elderly, which makes up a large population group. How old does it take to qualify for them to answer the questionnaire? It also needs attention.

Further, how many of those qualified for self-reporting can make appropriate judgments about their own well-being? Professor A. Deaton, the winner of the 2015 Nobel Prize in Economics, points out that in surveys on the overall life evaluation measures are far from perfect. People are often not sure what the questions in the questionnaire mean, and how they are expected to answer (Deaton, 2013). Professor Deaton himself worked at the World Bank and contributed to the Living Standards Measurement Surveys. In 1997, he published “The Analysis of household Surveys: A Micro-econometric Approach to Development Policy” (Deaton, 2013). As far as this type of surveys is concerned, he is a master both in theory and practice, so his comments and warnings are of great importance. Life evaluation measures are far from perfect...international comparisons can be compromised by the national differences in reporting styles (Deaton, 1997). Those of us who have been involved in constructing and critiquing these numbers should be much more skeptical and hesitant when using data.

There are differences in people's perception ability to distinguish. How to determine the number level of options? Is it a grouping of good, medium and poor enough, or a grade five or even seven? For sensitive people, exact level cannot be achieved with few grades; for insensitive people, too many grades are difficult for them to locate their answers. If a grading standard is used to tailor people with different sensitivities, the reliability of the measurement will be affected. Therefore, the questionnaire method also needs to assume that respondents' cognitive abilities are the same or their abilities are normally distributed, and the differences can be offset. What is the number of respondents and their distribution structure? Is it enough to meet the requirement of truthfully reaching an overall judgment?

The second is the moral hazard faced by the questionnaire method. To adopt the data results of the questionnaire method, it is necessary to assume that the respondents will perform according to their recognition ability and answer truthfully. However, the reality is that even if people can make accurate judgments about their own well-being status, why must they report it truthfully? It is precisely because respondents are conscious that they can make use of the questionnaire for their own benefit. Here we cannot ignore the reflexivity in the social phenomena emphasized by George Soros. If self-reporting infiltrate utilitarian factors, and civilians want more from the government and society to pretend to be unhappy, or fear that they will be suppressed to pretend to be happy, the results of the questionnaire may lead us astray.

How to eliminate various interference factors in this false self-reporting? This is one of the major games faced by the questionnaire method. The investigators need to grasp: Does the respondents answer the questions truthfully? What is the proportion of truthful answers? Can it support the research conclusion? Professor Deaton points out that many economists and philosophers have reservations about the validity and reliability of self-reported assessments. We do not always know what people are thinking when they answer these questions (Deaton, 2013).

Finally, the questionnaire method requires high quality of the measurers. Not only the quality of questionnaire design has a great impact on the survey results, but also the quality of questionnaire design in different regions should be roughly equivalent. Otherwise, the comparability of the measurement results cannot be guaranteed, which may mislead the analysis conclusions of the results. Professor Nick Wilkinson points out in Behavioral Economics that introspection of happiness may be guided by the situational context. Research indicates that people's evaluation of experiences is strongly affected by the prior questions they are asked. Such effects are known as Anchoring Effects (Wilkinson and Klaes, 2005). How to avoid this anchoring effect is the difficulty of questionnaire design, which is different from objective indicator recording, therefore, special attention is required.

3.3.2. Three limitations of the Willingness-to-Pay method

Willingness-to-Pay (WTP) is a common method in subjective well-being surveys. How much would you like to pay for a certain option? People's subjective well-being preferences are expressed by WTP. Welfare economics traditionally relies on the notion of WTP to extend the scope of the monetary measures to non-market aspects of life (Stiglitz et al., 2009a). However, the author believes that the WTP approach has three limitations.

The first limitation is the ad hoc effect. People usually give their WTP under the conceived situation, and they are not immersed in the scene. They are affected by the psychological differences between on-site and off-site, and it is impossible to know how much the impact is. Many people

always make a lot of money in simulated stock trading but lose money once they are in the actual stock market, which shows that the ad hoc effect is great.

The second limitation is the magnitude effect. The scale of the choice questions designed in the WTP questionnaire is often small. Due to the psychological survey object, a considerable part of the decisions can be made under the economic conditions of college students. Generally, the larger magnitude of a decision-making problem, the fewer people have ability of rational decision-making at this magnitude, and the worse the reliability of answers. The scale of real social issues varies. It is difficult to accurately cover people's real psychological desires with the WTP questionnaire.

The third limitation is the population income structure. For high-income earners, the marginal utility of payment is low, but the marginal utility of achieving a certain willingness is high. On the contrary, for low-income earners, the marginal utility of payment is high, while the marginal utility of achieving a certain willingness is not so high. Thus, the results of questionnaires often reflect the WTP of high-income people, which is an inevitable joint result of numerical average method (such as arithmetic average, geometric average and harmonic average). The position average method (such as median and mode) can avoid this deviation, but it is not easy to deal with mathematically, and it also faces the limitation of convenience of calculation in analysis.

How to reduce the negative impacts of these three limitations on the results of the questionnaire is the key that should be full paid attention to when using the WTP method. We should pay attention to the influence of these three restrictions when interpreting the connotation of the data obtained by the WTP method.

3.3.3. Representativeness of samples

Sampling survey is the main method of economic statistics. How to use this method in economic well-being measurement? Or under what conditions can this method be used? Or what problems should we pay attention to when using the sampling method? Can the sampling method that effective for physical production be copied and used to measure well-being? A very important aspect of the "whether or not" questions is the representativeness of the sample. The author divides it into two aspects.

The first is the representativeness of the well-being status and emotions of the reporting time to their overall well-being status and emotions for the reporting individuals. Individuals have different emotions at different times, different subjective feelings, moods and sorrows, may fluctuate greatly. The SSF Report suggests that it is best to get a report on personal feelings in a timely manner, which can reduce to a certain degree the bias caused by memory and social pressure (Stiglitz et al., 2010). However, it is impossible for respondents to keep recording and report all the time. After all, time being sacrificed for this could have been used to improve their lives. So, how often does the questionnaire survey need to be adjusted? Can the mood of the sample time represent all states of the reporter?

The second is the representativeness of the reporters to all subjects (sample population versus non-sample population). How many people need to be investigated to get an overall subjective well-being measurement? The whole population includes different types: adults, minors and the elderly, or the population suitable to answer the questionnaire and the population unsuitable to answer the questionnaire. Can the subjective well-being self-report of the investigated population represent subjective well-being of other types of population and the whole population? One thing is for sure that we cannot represent the subjective well-being self-evaluation of the so-called autistic patients. The representational bias must

exist. The problem lies in its size, and the process of adding up the micro-observation will transmit the representative bias to the macro-indicators. How to limit the reliability loss?

Because of the reverse effect of measurement resources, measurement frequency and sample size, the use of convenient samples is often unavoidable in the game, even for high-end researchers. In a groundbreaking study in 2010, Joseph Henrich, Steven J. Heine and Ara Norenzayan systematically surveyed all the papers published between 2003 and 2007 in the leading scientific journals under the six subfields in psychology. The study found that although the papers often claim about what the human mind is, most of them are based on exclusively samples of the WEIRD group. For example, in papers published in the *Journal of Personality and Social Psychology*, which should be arguably the most important journal in the subfield of social psychology, 96% of the samples belong to the WEIRD group (Western, educated, industrialized, rich, and democratic), and 68% are Americans. Moreover, 67% of American participants and 80% of non-American participants are psychology students. In other words, in all the papers published in this authoritative journal, more than two-thirds of the subjects of the experiment are psychology students in Western universities. The reason why psychology students participate in many of the studies is because their professors oblige them to (Harari, 2017). Those professors apparently know the methodological standards and requirements of subjective investigations, but use convenient samples for data analysis, which typically proves the practical difficulties of subjective investigations.

Psychological investigation has another major limitation: it is impossible to eliminate the influence of reflexivity. Even if we travel all over the world and study every community, we can still cover only a very limited segment of the *Homo Sapiens* psychological spectrum (Harari, 2017). Nowadays, everyone is affected by modernity, and we all are a member of the global village. The joke is that in the Kalahari Desert, the typical hunter-gatherer team consists of twenty hunters, twenty gatherers and fifty anthropologists. Psychologists pay attention to experiential investigation, where they observe the social reality at close range, and live, eat and work with surveyed objects. A total of forty people could be observed but fifty to observe. The presence of investigators is so strong that the behavior of surveyed objects might have changed.

Therefore, it is very important to establish a pattern of economic measurement in advance. So, is it empirical or false? Is it a matter of great concern, or a matter of humorless grin? How to maintain the professional ethics of economic measurers? A joke satirizes a Chinese journalist conducting a survey on the train during the Spring Festival, who asks passengers if they have bought tickets for the Spring Festival transport. This kind of “investigation” is problematic even with the minimum logic of thinking. When it comes to the limitations of subjective surveys, the error level of those American professors in social psychology are essentially comparable to that of this Chinese journalist.

3.3.4. Questioning the methodology of composite index

The notion of capabilities method and the fair allocation method need to take various objective factors affecting well-being as constituent indicators, and then calculate the composite indicator to get an evaluation of the overall economic well-being, such as the human development index and various happiness index.

The author points out two main flaws in the method of composite index in the “Multi-indicators Comprehensive Evaluation: Reflection on Methodology” (Qiu, 2018).

First, the correlation between the various component indicators may lead to the overlap of the information used, resulting in distortion of the synthesized information. The component indicators need to be related to the things being evaluated, but the correlation between the component indicators should be as small as possible. This is a contradictory requirement that cannot be fully met in economic reality.

In addition, there are concerns about insufficient basic information. Besides GNI (Gross National Income) and HDI (Human Development Index), only two representative indicators of education and health are selected. The remaining six features of objective factors affecting the quality of life mentioned above are not considered. Resources and environmental factors are not considered. It is not green (it has been proposed to build a green human development index). Can the integration of three factors represent human development? How to coordinate the selection of component indicators with the collection of basic information?

Second, there exists an issue of equivalent conversion in composite indicators. Once the composite formula is determined, a fixed equivalent conversion relationship between component indicators is determined. The author takes HDI as an example to illustrate that mathematical additivity is not equal to additivity in economic and social significance. The SSF Report gives a synthetic interpretation: adding the logarithm of per capita GDP to the level of life expectancy is equivalent to implicitly thinking that the value of an increase in life expectancy for Americans by one year is equal to twenty of the same increase for Indians (Stiglitz et al., 2010). As for the reason for this quantitative equivalence relationship, no one gives an explanation from the socio-economic sense.

The SSF Report also points out another flaw in the composite index method, that is, the concealment of structure changes by the average. The composite treatment ignores the correlation between the various well-being factors and does not reflect the state distribution within the economy. Even if the actual structure changes, as long as the average number of component indicators remains unchanged, the composite conclusion will remain unchanged.

The author's interpretation of this defect is that it means that that synthetic result does not have the ergodicity, that is, it cannot represent the various spatiotemporal states experienced by the changes of the things being evaluated, but it is only one of its many possible results. Composition regards part of state of the evaluated object as the whole state of the evaluated object. In other words, even if the distribution structure of various component indicators is different, as long as their average is the same, the same comprehensive evaluation conclusion can be reached. In this way, the comprehensive evaluation does not reduce the uncertainty that it attempts to reduce, or the data results of the comprehensive evaluation do not have the specific socio-economic implications that should have, and the comprehensive evaluation information is still uncertain. In view of this non-ergodicity, people should not make an absolute interpretation of the comprehensive evaluation results.

3.4. Text defects in the well-being measurement of the SSF Report

3.4.1. How to develop a corresponding indicator based on the concept of well-being measurement?

As far as the text is concerned, there are four major deficiencies in the well-being measurement:

First, there is no logical summary of the relationship between measurement concepts and measurement methods. Each philosophical concept has different measuring methods and different measuring mechanisms. The discussion in the section of SSF Report is too broad. How does the measurement logic run through the process from the concept of ability to the compilation of the human

development index, and from the abstract to the concrete? Is the calculation method based on this measurement concept unique? Does this logical relationship also exist in other methods? What is the reason for its existence? What is the difference between the methods? What are the different links? This section does not make a systematic discussion of the measuring mechanism, but pays close attention to logic of economic measurement, which is the key to improving the methodology of economic statistics.

Second, the SSF Report does not sufficiently summarize well-being measurement methods. In this section, there is no presentation and comparison of different economic well-being measurement methods, which is not as good as the discussion of the sustainable development measurement. The author will further analyze it in the “Sustainability of Sustainable Development Measurement”.

Third, the SSF Report does not provide a feasibility analysis and budget stress test for the improvement recommendations. The SSF Report points out that the measurement does not replace conventional economic indicators, but it provides an opportunity to enrich policy discussions and increase people’s awareness. The original well-being measurement is insufficient, and there are five additional aspects need to be included: to include subjective well-being in statistical surveys; to measure the influencing factors of well-being; to assess inequalities comprehensively; to assess the links between various quality-of-life domains; to construct composite index for users. These five points constitute the improvement recommendations of the expert committee.

In a nutshell, the solution to the problems reported by the SSF Report is "doing addition", among which at least the following issues need to be noticed:

First, the additional cost of “doing addition”. Taking the mutual influence of different well-being factors as an example, such as the impact of health and education on employment and economic security, the additional difficulty lies in that how much is the cost increase in different countries by adding a standard question to the regular survey? How much does the cost of survey increase if the classification is further refined? It is necessary to measure the cost of trial calculation, compare the cost-effectiveness of the measurement, and conduct a feasibility analysis. In these five aspects of economic well-being measurement, humanistic care needs to be implemented. After all, the public resources that society can use for measurement are limited, and it’s impossible to include all indicators.

Secondly, sort out the outstanding issues. There are still many unresolved issues in the measurement methods, which are worthy of discussion theoretically. The author advocates listing these issues as much as possible and clarifying the general ideas of solving different issues, so that it is possible to know whether the addition can be achieved.

The third is the difference between standard statistics and special statistics. A basic tenet of economic statistics is that the feasibility of small-scale special statistics differs greatly from that of official standard statistics, at least the budgets of the two statistics are very different. Where the former is feasible, the latter may not be feasible.

Finally, the tolerance of different levels of economic development to additional statistics. The measurement that the developed countries can conduct, the developing countries may not be able to do it. What are the requirements for incorporating well-being measures into standard statistics? How many countries have the conditions and what are there? Nowadays, many developing countries even have difficulty in routinizing GDP statistics. How to further develop economic well-being measurement? The SSF Report believes that not only in developing countries, it is necessary to go beyond the measurement of economic resources, where the traditional focus of much work on human development in the past, but also in the wealthy industrialized countries, the “Go Beyond” is of much

more significance. Where should the focus of well-being measurement be? The SSF Report has not been analyzed, so this issue needs to be followed up.

The SSF Report believes that recent advances in research have led to innovative and credible measures, some of which can be compared across countries in a reliable manner. Why is it? It remains to be further demonstrated. The SSF Report claims that these well-being measures, without replacing conventional economic indicators, have the potential to move from research to standard statistical practice. But there is no proof for this major conclusion, and no more detailed explanation, which is really worrying. What is the position of monetary aggregates method in economic well-being measurement?

The material standard of living is mainly measured by income, but income cannot cover other aspects of well-being. Income is a very important criterion for people to evaluate their living conditions, but it is not the only one. It cannot be said that being wealthy is happiness, nor can it be said that having money has nothing to do with happiness. The Equivalent Income Approach can be used to estimate non-cash items of economic well-being.

Due to the diversity of economic well-being itself, the notion of capabilities does not recognize economic rationality and its model, and monetary aggregates approach is weakened in the measurement of economic well-being, at least not highlighted in the section of quality of life.

The potential contradiction is that if in the measurement of current well-being (quality of life), monetary aggregates approach is difficult to be trusted, which is outside the three main well-being measurement conceptual approaches. Then how can the method of calibrating GDP stand? How can people accept it? Obviously, these two sections of the SSF Report need to be coordinated in this basic position.

3.4.2. Questioning measuring happiness by biotechnology

There are certain problems with the questionnaire method to measure subjective well-being. Some people advocate using biotechnology. The human brain can be implanted with a special chip to record the frequency of fluctuations in the nerves responsible for happiness, which can be used to measure individual happiness. In the future, biotechnology will be developed, and this concept can be fully realized. Yew-Kwang Ng, a well-known Chinese economist and a professor at Monash University in Australia, particularly agrees that biotechnology is more reliable than GDP statistics.

As we all know, with camera surveillance, human behavior will be different. With the chip implanted in the brain, knowing that “Big Brother Is Watching You”, every move will be restrained and the degree of freedom decreases, resulting in a decrease in happiness. This corresponds to the most basic characteristics of management, that is, management is anti-human. Therefore, the happiness value measured by biotechnology is the value after the original happiness is reduced. One additional item should be added to the true happiness value, which is the decrement in happiness brought about by the implanted chip. However, the trouble is that individuals have different degrees of sensitivity, and their happiness losses are different. How much should everyone’s happiness be added? A new measurement problem is brought forward.

In addition, is the biotechnology method a comprehensive measurement or a sample? If the child is implanted with a measuring chip as soon as he is born. If the child grows up and asks for the chip to be removed, is it allowed? With more people protesting, how should the society deal with it? If we respect their own wishes and convene volunteers to implant the measurement chips, can it meet the sample size requirement of the macro well-being measurement?

A more economical way is probably to simply remove the nerves responsible for pain. With the development of biological sciences, which nerve is responsible for happiness, anger, sadness and joy will be known. Then it can be done by an operation to remove the pain nerves at born. Without the need to measure at all, there is also no pain at all. But it's hard to imagine that if people have no pain, can happiness soar when good things happen? According to the past, people's feelings are relative, and the degree of suffering may be as high as the degree of happiness. If you can't feel any pain, then you can't feel much happiness neither. Is there any meaning in that kind of life? Behavioral economy refers to "No pain, no gain". Pain and happiness go hand in hand, how can human being resolve it?

Professor Deaton believes: Even if people carried a watch-like measuring instrument on their body that records every happy mood, there is no reason to assume that these data would be useful for assessing the happiness in their lives. Many different aspects are related to happiness. They are related to each other but never the same (Deaton, 2013).

3.5. The necessity of economic well-being measurement

3.5.1. The boundary issue of economic measurement as public goods

Regardless of the extension, economic well-being measurement belongs to a kind of public goods, so there must be a boundary issue, that is, its necessity ("if not"): to what extent should economic entities such as the government or non-governmental organizations intervene? As a component of the output of the government, where should the economic measurement end? Is it already complicated GDP statistics not enough?

Measuring economic well-being is naturally to improve people's livelihood. However, out of public conscience, it is not necessarily natural to have the legal rationality of behavior. If quality of life measurement itself causes a heavy burden on people's quality of life, if the production of public goods squeezes the private living space, the nature of things will be reversed, and cost-benefit analysis is required, which may result in the question of what is the necessity of well-being measurement.

There are two kinds of civil rights thoughts in the world. One is represented by the English proverb, "The storm may enter, the rain may enter, but the King of England cannot enter!" The house of common people was broken, so long as the doorframe was standing there, the king had to ask the permission of master of the house before he entered. Even the King should respect private rights. The other is represented by an ancient Chinese saying, "Every inch of land belongs to the king". Since it is the king's land, there is no limit to where the king can be. If we act in accordance with the former, economic well-being measurement should also fully consider respect for civil rights.

At a seminar on economic statistics, a professor of mathematical statistics once asked: is SNA a planned economy? Many people immediately informed that SNA is a market economy. The author thinks that the professor's economic intuition is good for raising this question. In fact, this question can be further explored. SNA is a product of the Second World War, that is, a product of the regulated economy. After the Second World War, the implementation of the "Marshall Plan" has been strengthened, and it does have a certain planned economy element. In general, there is no absolute boundary between market economy and planned economy. Measuring everything has a natural connection with planning everything and managing everything. Therefore, economic measurement needs to pay attention to the issue of moderation, especially economic well-being measurement.

3.5.2. Platform for economic well-being

In generally, income and wealth are the communication platform between individuals and the society, where individuals and families can move freely and can advance or retreat. What society should do is to create opportunities and environments for individuals to achieve happiness and to set up a well-being platform. As for which aspect of happiness everyone prefers, society should not interfere excessively. In a sense, this should also be an important economic measurement boundary.

How far should the economic well-being measure go? The so-called “Big Brother Is Watching You”, the warning in “Nineteen Eighty-Four” needs serious consideration. As taxpayers, the common people delegate public power to the government, but it is difficult to define what kind of public goods it provides, so government officials still have considerable discretion. According to the SSF Report, the economic well-being measurement should be specific to everyone. The first cross-cutting challenge for quality-of-life indicators is to detail the inequalities in individual conditions in the various dimensions of life, rather than just the average conditions in each country (Stiglitz et al., 2010). What should be questioned is: do people have to tell outsiders about their personal quality of life? Even in order to improve public well-being, must it be at the expense of exposing personal well-being status? Should the government measure my private life after paying the tax? Who should be the decision maker in the well-being measurement?

This involves the operation boundary of economic measurement. The author has specially discussed it in “The Boundary Paradox of Macro-measurement and Its Significance” (Qiu, 2012), which is the third measurement boundary. It concerns the balance between measuring relevance and measurement resource availability. It should not be mistaken to think that economic measures can be unlimited. It is good that the economic well-being measurement comes from humanistic concerns, but a sign of modern civilization is to maintain a proper sense of distance. Excessive humanistic concerns will infringe on personal privacy. If we insist on measuring economic well-being in depth, there may be conflicts between humanistic concerns and adherence to private space. Moreover, focusing on the actual effect of measurement, quite a few people may react negatively to economic well-being measurement, which will lead to the distortion of the data results, which involves the Neutral Paradox of Economic Measurement. A further analysis is made in “The Neutral Paradox of Economic Measurement and Its Significance” (to be published).

3.5.3. The policy significance of economic well-being measurement

Why should we measure economic well-being? Is it impossible to know how to build a happy society with the total production of society remaining unknown (assuming it is measurable)? Don't all sectors of society know how to pursue happiness? What is the significance of knowing the level of social well-being for the pursuit of happiness? In fact, what is more important is the true ability and opportunity to improve well-being. Bhutan's Gross National Happiness was once well-known all over the world and was highly praised. Although the economic level is not necessarily positively related to well-being, how can such a low level of economic and social development produce the well-being recognized by the people of the earth? One thing is clear. Most people cannot live in such a happy state for a long time.

Measurement means standard. Is it necessary to be uniformly guided to be happy? Happiness is contrary to uniformity. Biology wants diversity, and people need diversity even more. There was a

time when the kindergarten strictly controlled the children, and they had to sit neatly next to the wall, which is against nature and sad. The lesson was not far away. Why does happiness need a model? Do people have the right to steal happiness and escape compassion? The right to be alone, in a sense, the right to avoid social supervision, not to live under the camera, and to reduce the panic of being caught out at any time. Doesn't it take up the precious time that people should have leisure to fill in the well-being questionnaires and the Day Reconstruction Method" (DRM)? What if this time is really used to improve actual well-being? Taking the questionnaire reduce people's level of happiness. For the sake of public measurement, it requires people to recall things that should have been forgotten. Many studies have shown that hedonic introspection reduces people's face towards happiness, and those who are happy do not introspect too much. If so, isn't the subjective well-being questionnaires just the opposite of well-being?

3.5.4. The legitimacy of paying public resources for economic well-being measurement

The economic well-being questionnaire requires real-time reports to ensure the authenticity of the data results. To the extent that these feelings are reported in real time (Stiglitz et al., 2010). The SSF Report emphasizes that providing a long-term quality of life measurement that can be used for long-term monitoring at the individual level of each individual. In the author's understanding, this is to establish a set of systematic microeconomic statistics for well-being measurement.

The problem is that people cannot spend a lifetime consciously asking themselves how happy they are. In order to measure the total well-being, we need to measure the well-being of the individual, and pay part of the public resources for it. What is the legitimacy? There are two kinds of legitimacy: the legitimacy of the measurement itself and the legitimacy of consuming public resources for measuring expenditure.

As far as public resources for well-being measurement and personal information are concerned, there may be different combinations of willingness in different countries. How to make decisions? In developed countries, strong economic power and public resources are available, but personal privacy is unwilling to be monitored. Developing countries do not care much about privacy protection because of their poverty, but there is no better way to measure well-being with public resources than to use them directly to improve life.

If the recommendations for improving the well-being measurement in the SSF Report are implemented, society, especially poor countries, will be overburdened. Think about why a census can't be carried out every year, even if people all over the world go to register and do nothing else, the census does not work well, and is overburdened.

3.5.5. Mandatory information tax

Tax can be levied in non-monetary ways. What people often neglect in daily life is that we have been paying information tax all the time. And the more the era of big data, the heavier the information tax.

In a broad sense, the author asserts that anyone is related to statistics in some way, either statistics producer, or statistics user, or at least statistics object. Everyone is inside the bureau of statistics.

The meticulous and frequent questionnaire survey is a kind of heavy tax in society. Some people do not accept statistical surveys, but in fact, they cannot afford to hide. As soon as you turn on your

smart phone, you're already under statistics. Personal data such as Alipay and Drop-by-drop taxi and so on are involuntarily uploaded to the cloud. People are always greedy, and the cost of convenience is to be counted. In real life, we pay a lot for convenience, not only money, but also personal information, and even privacy. However, most people do not consciously pay for this kind of privacy information.

If we are fully aware of the various apparent and potential information costs, is well-being measurement still necessary?

3.5.6. Possibility of limited well-being measurement

What is the use of GDP in well-being measurement? Professor N. Gregory Mankiw believes that we can conclude that GDP is a good measure of economic wellbeing in most cases, but not all cases, purposes. It is important to keep in mind what GDP includes and what it leaves out (Mankiw, 2015). If Mankiw's conclusion holds, is it necessary to start anew to develop economic well-being measurement?

"Understanding National Accounts" puts forward a thought: better welfare measures within the national accounts. GDP is only one indicator in a mature account system. Economic welfare indicators that are better than GDP do exist, that is, those measures for households (rather than to the economy as a whole), as individuals and households are the natural basis for evaluating well-being (Lequiller and Blades, 2014).

In SNA, the logical relationship of indicators from production to final use is that GDP adjusts net income of domestic and foreign factors to get Gross National Income (GNI), subtracts consumption of Fixed Capital to get Net National Income (NNI), and then subtracts tax to get Net Disposable Income (NDI), and plus FiK to get Adjusted Net Disposable Income (ANDI). NDI subtracts Savings is the Final Consumption Expenditure of households, while ANDI subtracts Savings is the AFC of households.

FCE, especially AFC, can be used as a platform to measure economic well-being. Further adjustments are needed: First, eliminate the impact of population size. Second, consider the factors of income distribution and inequality, involving micro-data of different types of households. Third, include the stock factor, because savings (negative savings) are closely related to the current economic well-being.

Continuation the ideas of OECD experts, Francois Lequiller and Derek Blades, can we consider adopting the economic well-being measurement model of the "standard statistics + special survey"?—Existing income and consumption indicators of SNA households are adjusted as the content of standard statistics, while other relevant well-being information is collected through special surveys.

The discussion of the necessity of measuring economic well-being should not be totally negated, but at least help limit it to a certain extent. Doing addition is not the only way out, and it is not necessarily the best choice to start anew.

4. Sustainability of sustainable development measurement

4.1. Reflections on the premises of the concept of sustainable development

Many people believe that sustainable development is justified, and it is the general trend. In fact, there are still many fundamental issues to be considered.

4.1.1. Sustainability of whom

Most people tend to regard sustainable development as a whole concept, which aims at comparing contemporary people with future generations. In economic reality, there are not only inter-generational differences, but also intra-generational differences. What is the relationship between the two? Which is more important? People's perception of well-being comes from time and space comparison. Therefore, people are more likely to notice the intra-generational differences, while the impact of the inter-generational differences on people is probable and relatively unnoticeable, at least contemporary people have opportunity to shirk their responsibilities.

Intra-generational differences make structural issues prominent: sustainability of whom? Does the well-being sustainability of the wealthy second generation take precedence over the current well-being of the poor? People have different social status, different levels of demand, different preference focuses and different arbitrage goals. As a result, price elasticities for sustainable development of different actors are also different, and prices of sustainability are also different. Inter-generational sustainability is a necessity for the rich but a luxury for the poor. Only when the gap between the rich and the poor within intra-generation is less than a certain level can inter-generational sustainability be truly inclusive and can be truly promoted.

From the beginning, the contradiction between intergenerational sustainability and intra-generational differences existed. In the late 1960s, some visionary elites in developed countries began to advocate sustainable development and prepare for a large-scale international environmental conference in Stockholm, Sweden. Pakistani economist and founder of the Human Development Index, Mahbub UI Haq, once believed that the emphasis on environmental issues might prevent the industrialization of developing countries at the time, so poor countries should not be forced to accept the provisions formulated by industrialized countries.

The contradictions between the macro and the micro, between public goods and private goods, between universal values and national interests always exist, and the problems of hierarchy, structure, or quality difference always exist. It is difficult to achieve promoting sustainable development by promoting universal values alone. What exactly is the "smart power" of the European and American elites? The author's interpretation is that in a situation where national interests and universal values have a relatively large intersection, such countries can more fully safeguard their national interests by pursuing universal values. That is, power that can make its national interests and universal values intersect is the smart power, just like the enterprise that can be the industry standard is the high-end enterprise. If we really start from universal values, certain actions of developed countries will lose their occupation of the moral high ground. Rich countries accuse developing countries of wasting resources and polluting the environment. Can developed countries take action to stop exporting polluting industries and domestic waste to poor countries?

From the perspective of resource allocation, poverty today and the plight of future generations are mutually cause and effect. Resources are inherently limited. If we pay too much attention to intergenerational differences and are biased towards future generations, it will have a negative impact on the gap between the rich and the poor within the generations. If the contemporary poverty issue is still serious, but the resources are reserved for the future generations, what is the rationality or legitimacy? If the difference between the rich and the poor within a generation causes social stagnation or even turmoil, what is the basis for inter-generational sustainable development?

It should be recognized that objectively there exists the issue of sharing responsibilities between rich and poor countries. It requires that poor countries and rich countries to set a standard for sustainable development, which is advanced for poor countries to even share responsibilities for rich countries. From the perspective of history, the present and the future, responsibilities of resources and environment should not be shared equally. Generally, there are four reasons: First, the historical plundering of debts by the great powers. Second, the level of demand for sustainability. Third, the moral high ground of the rich. Fourth, the different economic capacity and wealth stock.

Therefore, rich countries should pay more for environmental protection (Qiu and Chen, 2007). When the global sustainable development goals were formulated, rich countries reluctantly agreed to subsidize the limited funds of poor countries, and most of them are not in place. Between national interests and universal values, the priorities of developed countries are clear. President Donald Trump's withdrawal from the Paris climate agreement is a typical example.

4.1.2. Stage burden of historical debts

A process is needed to achieve sustainable development. Environmental and resource debts are caused by the accumulation of several generations. It is neither possible nor reasonable to achieve sustainable development in the contemporary era. Environmental protection debts should be gradually compensated in several generations. In 1999, the author put forward the Hierarchy of Sustainable Development. Different countries and different stages of development should have different requirements, one-size-fits-all is not feasible (Qiu and Song, 1999).

It should be noted that people attach different importance to sustainable development in different economic periods. Only under the optimistic background of Francis Fukuyama's "The End of History and the Last Man", can sustainable development get more recognition. If there is an underestimation or crisis in the world economy, people will inevitably give more consideration to current interests.

4.1.3. New debt for environmental resources caused by sustainable development

There is no free lunch in the world. Sustainable development also requires a price. Cleaner production is only a relative concept. Clean energy production is good, but raw materials, spare parts or by-products may be highly polluting. How to compare the cost-benefit? How to choose? For example, how to dispose of the trichlorosilane and silicon tetrachloride brought about by the production of solar cells? How to consider the potential impact of "Not in My Backyard" (NIMBY) in national competition?

Paying attention to "cleaner production" cannot be absolute. The society has not developed to that level, where clean production all over the world is enough to make the life of all mankind run smoothly. Cleaner production is only a dynamic concept in development. In other words, the so-called "non-cleaner production" still needs to exist, or partially exists, or exists for a while but is gradually reducing. "Non-cleaner production" is often a forced job for the poor, no matter how the price is displayed. It is by no means their preference, but it is the invisible foundation of the elegant life of the nobility. Therefore, the developed countries should not give up the responsibility of resources and the environment. The nobles should not accuse the cook of cruelty when eating roast pork.

4.1.4. Same preference assumption and same Intelligence assumption

After careful deliberation, the concept of sustainable development implies two basic assumptions.

One is the assumption of the same preference, that is, the preferences of future generations are the same as ours. We like oil, and so do our children and grandchildren.

The other assumption is the same intelligence assumption, that is, the intelligence quotient and the ability of future generations are the same as ours, and they can only follow the path we set. What is the likelihood of these two basic assumptions? Is there no progress in history? There may be new alternative sources of energy, and the scarcity of oil will be greatly reduced. The differences in preferences and intelligence of the same generation are so big, and how can the preference and intelligence of different generations be the same?

From the perspective of moral high ground, we can criticize the use of resources as “eating the ancestor’s food and cutting off the food for the children and grandchildren”. But objectively summing up the human genetic experience, a moderate shortage may be conducive to breaking “well-being curse”, so there is another conclusion: “children and grandchildren can take care of themselves when they grow up, so parents don’t have to work too hard for their future”. We can’t say which of these two views is 100% correct. How much resources for stock and how much for use? There is no consensus.

4.2. *Reflections on the concept premises of sustainable development measurement*

The concept premise of sustainable development measurement comes from a correct understanding of sustainable development, or from a general understanding of economic measurement.

4.2.1. Sustainable development measurement is a generalized measurement

In a narrow sense, the physical measurement tradition emphasizes the ex-post accounting. Sustainable development measurement is a contradictory concept. Sustainable development is incompatible with measurement. Whether something is sustainable or not can only be determined after experience. Understanding sustainable development requires the use of generalized measurement concept, which must include estimation and prediction components. This needs to be extended from a narrow measure to a generalized measure.

“A Brief History of Tomorrow” has broken the famine of history. It has not yet experienced and has become “history”! Time is continuous, and “today” is transient. There is no absolute boundary between the past and the future, but the time frame made by human beings.

4.2.2. The spatial scope of sustainable development measurement

What is the spatial scope of sustainable development measurement? National or global? Do you also consider outer space? Where does the outer space stop? What is the feasibility of planetary nomadism way of life? As far as the earth itself is concerned, there is also a problem of the depth of resource development. If new energy sources such as combustible ice can replace oil, how significant is the sustainability of oil reserves? Seawater desalination technology has solved the crisis of

freshwater shortage. If freshwater shortage reaches a certain extent, the use of this technology may be not necessarily expensive. Considering the present condition, how much is the comprehensive cost of the South-to-North Water Diversion Project and the West-to-East Natural Gas Transmission Project? Compared with the use of seawater desalination in northern coastal areas, which one is more expensive? It can also be demonstrated in depth and lessons can be learned.

The warming of the earth must be phased and fluctuating. The key lies in the duration and the amplitude of fluctuations, and whether it is beyond the tolerance of human beings. The interaction between human beings and nature is extremely asymmetrical. Does the so-called global warming phenomenon really exist? Is there any possibility to enter the ice age in near future? If global warming is conclusive, then what is the reason? Is it the climate cycle of the universe or the result of human activities? Different scientists have different opinions on these issues. This matter cannot be concluded by a simple majority opinion. In addition, more attention should be paid to the hidden attempts of competition among countries behind the opinions. In a word, the basic cognition of sustainable development is still in doubt, and the measurement conclusions of sustainable development in different spatial scopes are different.

4.2.3. Time range of sustainable development measurement

How long does time range of sustainable development measurement last? Permanent sustainable development, or 100 years? Keynes said that in the long run we are all dead (Ehsan, 2016), so he focuses on short-term analysis.

Suppose an extreme situation in which human beings will be destroyed with a higher probability within a foreseeable period. Is it still meaningful to adhere to sustainable development? Of course, the most basic hypothesis implicit in individual behavior is “30,000-day hypothesis” of survival, similar to the “going-concern assumption” of enterprises, which is time-based concept of sustainable development.

In addition, the old debts of resources and environment mentioned above needs to be burdened by stages, which is an internal structure issue in the division of its time range. The uncertainty lies in: How many generations will the repayment be? If linear decline occurs, how long is the period determined? If a non-linearity decrease is performed, will the pattern be first more, then less, or first less, then more? Can contingency be achieved? These difficulties are similar to the difficulties in the fixed capital consumption measurement, which requires at least two premises: cycle length and consumption rate.

Discussing the scope of space and time is to determine the space-time framework of sustainable development measurement, which is the premise of establishing the indicators and measurement methods.

4.2.4. Difficulties in stock measurement becomes basic difficulties in sustainable development measurement

At different stages of human development, the importance of stock measurement and flow measurement varies. When William Petty wrote “*Political Arithmetick*”, he focused on stock measurement. Later, economic statistics was the parallel measurement of stock and flow. Modern economic statistics focused on income flow analysis, that is, mainly flow measurement. Since the 21st century, the new economy has flourished, and social well-being measurement has been emphasized. Stock accounting has once again risen to a very important position.

The SSF Report points out that the measurement of sustainable development is mainly a stock-based, capital-based or wealth-based approach to sustainability. Various forms of capital constitute generalized social capital, that is, total social capital. However, it is not easy to define it in connotation. The four sub-components are relatively clear.

First, social capital, in a narrow sense, including economic security, personal security, political participation and communication, can be listed as its main items at present, but many intangible contents are difficult to define and quantify. The SSF Report even argues that this part of content is too uncertain to be included in the conventional measures.

Second, human capital measurement is still controversial and has not been included in international standards. If measured according to the concept of human capital, there will be no Final Consumption item, which should be expressed as intermediate consumption of human capital, which means that the accounting concepts of different measures are contradictory. The measurement logic of physical capital and human capital is not coherent.

Third, resource capital includes environmental capital. If it is necessary to measure the total amount and add it to other capitals, there are only two methods: monetary aggregate and composite index. The most difficult problem of the former is its pricing problem, while the latter cannot escape barriers of Equivalent Conversion. In conclusion, both are the problem of weight determination, that is, the cognitive problem of the structure, and the core difficulty of economic measurement.

Forth, compared with the first three kinds of capital, physical capital measurement is slightly easier. But compared with economic flow measurement, such as GDP statistics, stock measurement is much more difficult.

The crux of the problem is that if there are major difficulties in capital measurement as a basic concept, then it is difficult to fully establish sustainable development measurement on it.

4.2.5. To measure sustainable development humbly

Economic measurement should first make clear object (here is sustainable development), which is the need of comparing quality. The measurement and comparison of the real economy often need to take the country as the basic unit, but economic resources can be imported and exported, and environmental pollution can also be imported and exported, so can contributions of different countries be measured? How to measure it? If there is no scientific method to measure such contributions clearly, country should not be used as the basic unit of measurement and comparison. From an organizational point of view, it should be regarded as a basic unit, but it is difficult to become a competent basic unit, which means that the inherent contradictions in sustainable development measurement are difficult to properly solve.

Why discuss the sustainability of sustainable development measurement? Operationally, compared with monetary aggregates, dimensionless synthesis seems to be far away from the socio-economic reality, so the focus is mainly on the equivalence treatment of externalities, pricing of resources and environment, or how to determine approximate equivalence between different factors. More generally, how to determine the indicator weight? This alone is enough to make sustainable development and its measurement difficult.

If the premise of these measurement concepts cannot be solved, the so-called sustainable development measurement, making projects, making observations and making inferences can only be false evidence analysis, which is specious and does not have great socio-economic significance. No

matter how profound the mathematical method used, if there is no real integration with economic reality, it is just playing with technology, nothing more than a digital game. Some people have mastered some quantitative analysis methods, thinking that they can be used everywhere, regardless of the peculiarities of economic reality, and confuse numerical calculation with economic measurability. Seemingly professional, but actually layman. The measurement logic is not really connected, the method is not supported by methodology, and the measurement is unsustainable.

Human beings should be humble in the face of nature and destiny. Sustainable development is a manifestation of human humility. The same should be true for sustainable development measurement. Don't mistakenly think that human beings are able to measure sustainable development. Although human beings have been exploring for more than 30 years, there has been no substantial progress. Therefore, although it is difficult to get rid of the predicament of conceptual premises, it still requires in-depth thinking.

4.3. Criticism of the four kinds of measuring approaches of sustainable development

Sustainable development measurement began in the 1990s, and people used various indicators to measure sustainability. Providing a brief summary of the very abundant literature that have been devoted to the measurement of sustainability or durable development is not an easy task. We will use an imperfect but simple typology that distinguishes (i) large and eclectic dashboards, (ii) composite index, (iii) index that consists of correcting GDP in a more or less extensive way and (iv) index that essentially focuses on measuring the degree of excessive consumption of our resources (Stiglitz et al., 2010). The fourth category is called the index of special mention resources by the author. In 1990, the author made a doctoral dissertation to summarize categories of indicators and methods. There are physical indicator, monetary aggregate, composite index, and multi-indicators comprehensive evaluation. The two generalizations fit together.

4.3.1. Large dashboards or sets of indicators

The dashboard is a vivid statement, which is easier to understand in the automobile culture. In fact, it is the indicator system emphasized in the Oriental Accounting System (i.e., MPS). This approach involves gathering and ordering a series of indicators that bear a direct or indirect relationship to the socio-economic progress and its durability (Stiglitz et al., 2010). It is an indicator method of grouping and hierarchical processing. The effect of the dashboard approach is that it is an initial step in any analysis of sustainability.

There is a characteristic in the use of indicator system method: people usually take a supportive attitude towards whether to establish an indicator system or not, but often have different ideas about its indicator composition. People are pursuing comprehensiveness, and the constructors of the indicator system often compromise with various advisory opinions and suggestions, thus a pattern of "preferring too much rather than too little" has been formed. The reason also lies in people's rational choices: overused indicators increase the cost of measurement at most, and too little indicators may be a manifestation of poor professionalism. Therefore, the usual problem with dashboards is that the number of indicators is increasing. Only when there are too many to the extent that everyone feels that there is too much content, can it be reduced when revised, and it may increase in the future, as is

the case with the basic development model. Therefore, to interpret the dashboard, one should hold the Occam's Razor and grasp the key points.

The construction of the indicator system is related to the distinction between weak sustainability and strong sustainability, which is also a content of the conceptual premise. Weak sustainability reconciliation allows comprehensive consideration, and the various dimensions are relatively equal. Strong synthesizability of all constituent indicators, and the value of indicators can compensate for each other. The weak approach to sustainability considers that good performance in some dimensions can compensate for poor performance in others. This allows a global assessment of sustainability using mono-dimensional index. Strong sustainability advocates setting a baseline, and it will be rejected by one vote if it is below the baseline. The comprehensiveness of the assessment is relatively poor. The strong approach argues that sustainability requires separately maintaining quantity or quality of many different environmental items. Following this up, therefore, it requires large sets of separate statistics, each pertaining to one sub-domain of global sustainability. Both have advantages and disadvantages. It does not mean that which standard is necessarily superior. Social preference is closely related to the choice of measurement perspective.

Dashboards nevertheless suffer because of their heterogeneity, at least in the case of very large and eclectic ones, and most lack indications about causal links, their relationship to sustainability, and/or the hierarchies amongst the indicators used. Further, as communication instruments, one frequent criticism is that they lack what has made GDP a success: the powerful attraction of a single headline figure allowing simple comparisons of socioeconomic performance over time or across countries.

To make decisions and make choices, we need a single headline figure. To make overall comparisons between specific time and space, we need to add up, and we need information results that facilitate action. This leads to the use of composite index and monetary aggregates.

4.3.2. Composite index for sustainable development

Some scholars believe that composite indicators are superior to monetary aggregates. They do not believe in price signals. Because composite indicators do not involve any form of market valuation, they believe that this aggregation method is superior to monetary aggregation. The compromise method synthesizes different viewpoints, which is another reason why it is popular.

The composite indicators for measuring sustainable development mainly include: Human Development Index (HDI), Index of Social Health (ISH) and Index of Economic Well-being (IEWB), developed in the 1990s; Happy Life Index (HLI), Sustainable Social Index (SSI) and Environmental Performance Index (EPI), developed in the first decade of the 21st century; Happy Income Index (HII), Happy Planet Index (HPI) and Better Life Index (BLI) developed in the second decade of the 21st century.

Index and indicators abroad are indistinguishable in many cases, and many people mix them up. What does it mean on special occasion? It needs to be clearly distinguished.

In fact, the index essentially informs us about the comprehensive situation of environmental quality, the pressure on resources and the intensity of environmental policy, but not about whether a country is actually on a sustainable path: no threshold value can be defined on either side of which we would be able to say that a country is or is not on a sustainable path. As emphasized in the SSF Report, it should be able to judge a country's development path: is it sustainable or unsustainable? The author thinks this measurement goal is difficult to achieve, and it is impossible to reach a

consensus to define a threshold. The threshold level is related to the spatial scope. Resources can be imported and exported, and the environment interacts between countries. It is hard to say whether a country or a region is sustainable. How large is the spatial scope of the sustainability measure mentioned above? If the scope is too large, it may exceed the measurement capability. If the scope is too small, there may be a spatial interaction that cannot be clearly cut and measured.

There are two main points in the operation of composite index. One is the linear scaling mentioned in the SSF Report, that is, proportional estimation, which is one of the standardized methods. However, the first issue is the clear definition of sustainable development for each component indicator. We should examine each component behind the general indicator more carefully and pay attention to its socio-economic significance. Another is the process of weighting: whether it is GDP statistics or composite index. The problem is not that these weighting processes are hidden, non-transparent or non-replicable. This has something to do with issue of Equivalent Conversion emphasized by the author, but it is not exactly the same.

Different variables affect the overall sustainability. How to understand the reasons for assigning different relative values to them? This is the Equivalent Conversion. In other words, mathematical additivity cannot automatically guarantee the additivity of socio-economic significance. It is not enough to add mathematically, but also to pay attention to its additivity in the socio-economic significance. The SSF Report points out that the normative implication of composite indicators is rarely clearly explained and proven, which confirms the author's point of view. It is precisely the key to this composite that no one has made clear. The author believes that this is one of the major measurement problems faced by mankind, and it is difficult to give a satisfactory answer. Some scholars may be aware of this but choose to keep it secret, while some scholars have not even thought about this level.

The author believes that the normative definition of composite indicators cannot be clearly stated and justified before comprehensive evaluation, and the implicit contradictions cannot be dealt with at least for now. Because a certain equivalent conversion relationship must be related to a specific space-time, it is impossible to generalize this equivalent conversion relationship in advance. The key difficulty is, where does the selected equivalent conversion relationship come from? Is it the experience of European and American countries, or the experience of developing countries? Composite indicators need a unified measurement framework for spatial comparison. However, the determination of the relative value of different evaluations should be based on the relative relationship of the evaluated objects in a specific space-time. This is where its unity and specificity requirements contradict. In other words, the general equivalent conversion relationship cannot ergodic various special occasions of the evaluated object. The lack of ergodicity means that we regard a certain special as a general, which is a major hidden risk of economic measurement and empirical analysis.

4.3.3. Adjusted GDPs

Because of the difficulty in explaining composite indicators in the socio-economic significance, some scholars advocate those monetary aggregates method should be adopted to measure sustainable development, which is expressed as Adjusted GDPs.

The idea of Adjusted GDPs

The proposition of Adjusted GDPs implies the understanding that the SNA accounting paradigm is one of the greatest inventions of mankind in the 20th century, “Beyond GDP” is not so easy. At most, it is possible to achieve “GDP and Beyond”. Note that these two propositions are quite different. One excludes GDP and the other is based on GDP. According to the scholars, composite indicators are even more unreliable, and monetary aggregates must be used. These scholars also acknowledge that GDP is flawed, but they can make appropriate adjustments to it. Incorporating sustainability-related factors into GDP is no more than two aspects. Adding factors that are conducive to sustainability and subtracting factors that are not conducive to sustainability. The so-called adjustment means doing a bit of addition and subtraction, which is simple in calculation but difficult in judgment.

The SSF Report points out that there is a more fundamental problem with Green GDP, which also applies to Nordhaus and Tobin’s SMEW and to the ISEW/GNI index. None of these measures characterize sustainability. What we ultimately need is an assessment of how far we are from these sustainable targets.

According to the definition of sustainable development, it should be a measure of overconsumption or underinvestment. This conceptual premise is very important for the selection and use of Adjusted GDPs. The approach of Adjusted GDPs can be summarized as the Equivalent Income Method, which belongs to the category of accounting/monetary method. It contains a series of indicators: Measure of Economic Welfare (MEW), Sustainable Measure of Economic Welfare (SMEW), Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), Green GDP, Genuine Saving (GS), Adjusted Net Savings (ANS), and so on.

Seminal contribution of Tobin and Nordhaus

James Tobin and William Nordhaus are the originators of the idea of Adjusted GDPs. In their seminal paper in 1973, they questioned the limitation of GDP on welfare measurement. They not only put forward the concept of Regrettable Expenses (Defensive Expenditure), but also proposed how to improve adjustment thinking and to build the Measure of Economic Welfare (MEW) obtained by subtracting from total private consumption a number of components that do not contribute positively to welfare (such as commuting or legal services) and by adding monetary estimates of activities that contribute positively to welfare (such as leisure or work at home), and then further built a Sustainable Measure of Economic Welfare (SMEW).

Factors that do not contribute positively to welfare (such as commuting or legal services), Tobin and Nordhaus argued that such items have no positive impact on welfare, and the author questions this. Why do people commute? Not just to work for money, try to compare two specific situations: two hours less off-duty leisure time per day for commuting, but living in a high-quality environment; two hours more leisure time for avoiding commuting, living near the workplace but only in a low-quality environment. One is high quality of living environment but little leisure time, while the other is long leisure time but low quality of living environment. Commuting is a means of choosing between different well-being schemes. Commuting is totally regarded as paying for production of the enterprises, which is somewhat one-sided. As Nick Wilkinson points out in “*An Introduction to Behavioral Economics*”, they also endure lengthy commutes. Who cares about a couple of hours a day in a car when you have a McMansion to come home to (Wilkinson and Klaes, 2012)?

In addition, there are other influencing factors related to well-being, such as what are the conditions of commuting: Is it commuting in rich countries or in poor countries? Are there any housing options near the workplace? Can individuals afford the housing prices near their workplace? What is the difference in living conditions and environments in different locations? Is commuting a revealed preference or not an option? When adjusting GDP, we need to think more about it.

It is also doubtful to assert that legal services have no positive impact on well-being unless the legal provisions are particularly cumbersome and confusing, at which time legal services can only offset the negative effects professional barriers on the people. However, it cannot be assumed that people live in an environment where the law takes effect automatically. Only under this hypothetical premise can legal services not increase well-being.

If the author's doubts are valid, it means that the so-called subtracted items affecting well-being may not be all deductions, and the added items may not be all additions. Direction of GDP correction (adjustment) by a single economic activity may be bi-directional. However, once the correction (adjustment) requires bi-directional consideration, how much should be subtracted? How much should be added? More controversy will arise, and there will be no ultimate arbiter. Addition and subtraction are very easy to calculate, but why to add and why to subtract, and how to determine its magnitude are often controversial. With a kind of fuzzy uncertainty, it is impossible to apply a fixed model.

The implicit premise of the correction (adjustment) method is that the measurer needs to make value judgments on various economic behaviors: what kind of factors affect the increase or decrease of economic well-being? If the spatial and temporal pattern of measurement needs to be expanded, it will be more difficult to reach a consensus. Normative issues also involve the "Neutral Antinomy of Economic Measurement". The author has made a more systematic discussion in the paper "*Neutral Antinomy of Economic Measurement and Its Significance*" (to be published).

Additionally, based upon their MEW, they built a Sustainable of Economic Welfare (SMEW) considering changes in total wealth. To convert the MEW into the SMEW, Nordhaus and Tobin used an estimate of total public and private wealth including Reproducible Capital, Non-Reproducible Capital (limited to land and net foreign assets), Educational Capital (based on the cumulated cost of years spent on education by people belonging to the labor force) and Health Capital, based on the permanent inventory method with a depreciation rate of 20% per year (Stiglitz et al., 2010). But why takes 20% instead of 25% or other values is the question that the SSF Report does not specify. The MEW still has a flaw: it does not include factors of environmental damage and natural resource depletion.

The development of the measurement of economic welfare method

Two schools of economic welfare measurement methods emerged. One has tried to enrich Nordhaus and Tobin's approach, which sometimes deviates increasingly from the criterion of accounting consistency. Examples include the Index of Sustainable Economic Welfare (ISEW) and the Genuine Progress Indicator (GPI). These indicators make up for the flaws of the MEW by deducting the cost estimates of water, air and noise pollution from consumption, while considering the loss of wetlands, farmland, and primary forests, the depletion of other natural resources, and the destruction of carbon dioxide. The depletion of natural resources is valued by the investment necessary to generate equivalent renewable alternatives.

The other strand is more firmly integrated into the realm of national accounting. It is based on the so-called System of integrated Environmental Economic Accounts (SEEA), as a satellite account of the Standard National Accounts. The SEEA brings together economic and environmental

information in a common framework to measure the contribution of the environment to the economy and the impact of the economy on the environment. The current macro-accounting model is “SNA Core Account+”. The goal of this strand is to strive to mainstream environmental and resource accounting. However, after so many years, the SEEA has not yet been integrated into the core account of the SNA. Obviously, the measurement logic of the SNA and the SEEA still has major obstacles that cannot be connected and accommodated.

The SEEA consists of four categories of accounts. The first considers purely physical data related to the flow of materials (materials drawn into the economy and residuals produced as waste) and energy. The second category takes those elements of the existing SNA that are relevant to the goods management of the environment and makes the environment-related transactions more explicit. The third category measures environmental assets in the form of physical indicators and currency (timber stock accounts, for instance). These first three categories of SEEA are vital building blocks for any form of sustainability indicators. But what is at stake here is the fourth and last category of SEEA, which is to adjust the existing SNA to include the impact of the economy on the environment in the accounting, in which three adjustment items are considered: one is Resource Consumption, one is Defensive Expenditures, and the other is Environmental Degradation.

Green GDP

It is these environmental adjustments to the existing SNA aggregates that are better known under the rather loose expression of Green GDP, which is an extension of the concept of net domestic product. Indeed, just as GDP (Gross) is converted into NDP (Net) by deducting the consumption of Fixed Assets (depreciation of produced capital), the idea of Green GDP is that it would be meaningful to compute an “EA-NDP” (environmentally-adjusted) that takes into account the consumption of natural resources. The latter would comprise the depletion of resources (the over-use of environmental assets as inputs of the production process) and the degradation of the environment (the value of the decline in the quality of a certain resource, roughly speaking). The concept of Green GDP is not complicated. It is just doing a bit of addition and subtraction based on conventional GDP. In summary, there are three main types of GDP: one is conventional, one is green, and the other is black, which is the added value of the underground economy.

The SSF Report evaluates this approach that valuing environmental inputs into the economic system is the (relatively) easier step. Since these inputs are incorporated into products sold on the market, it is possible (in principle) to directly assign a value based on market principles. In contrast, there is no direct way to assign a value to pollution emissions, as they are outputs. All the indirect methods of valuation will, to some extent, depend on “if...will...” scenarios, which means it goes beyond the scope of post-accounting, has a strong speculative nature, and is in a hypothetical situation.

The first part of this paper emphasizes the conceptual premise, and its role is reflected in this. When discussing GDP statistics, the author proposes that the Observed Value Priority Principle should be followed, and the basic data should be mainly based on observations. If item adjustments are needed, what are the key points of adjustment? How much adjustment should be made? When adjusted too much, the proportion of observations will decrease, which may conflict with the Observed Value Priority Principle. What is the priority of observations? What proportion of observations should be? What proportion should the estimated values account for? It is impossible to give a clear explanation in advance, although we should work in this direction.

The US President Bill Clinton issued a presidential decree in 1993, requiring the US Bureau of Economic Analysis (BEA) to release Green GDP data every year. The data has not been released for so many years, even the United States is the most economically developed country in the world, which shows that Green GDP is indeed not that simple.

There is a way to explain this, which is the putting to extreme and testing fallacy summarized by the author: push things to the extreme and know whether they exist. Green GDP is nothing more than an added item or subtracted item of conventional GDP. What to subtract from? Environmental damage and resource depletion. Imagine, if the cost of resource depletion and the price of resources are enlarged, subtracted item is greater than regular GDP, and Green GDP is negative. How should we deal with it? If we insist on green development, it seems that we could only stop production.

Some people over-advocate green development and advocate zero growth and zero emissions. In fact, this goal cannot be achieved. People live in the world will consume sources. Everyone is a garbage maker. The core controversy is: how to determine the price of resources and environment? is the adjusted green GDP reliable?

Adjusted Net Savings (ANS)

GDP usually represents consumption and investment from the perspective of flow, which involves sustainable development. The connection between flow and stock needs to be considered. SNA describes the economic process with five major subsystems, one stock subsystem and four flow subsystems. There is a corresponding relationship between the two. The concept of sustainable development emphasizes what we leave behind for future generations, that is, stock. The SSF Report believes that though such indicators tend to be presented in flow terms, they are built upon the assumption that some stocks that are relevant for sustainability correspond to the measured flows. Adjusted Net Savings (also known as Genuine Savings or Genuine Investment) is a sustainability indicator that builds on the concepts of green national accounts but is recreated from the perspective of stock and wealth rather than flows of income or consumption.

The theoretical background is the idea that sustainability requires the maintenance of a stable stock of Extended Wealth, which is not limited to natural resources but also includes physical, productive capital, as measured in traditional National Accounts, and Human Capital. Net Adjusted Savings is defined as the change in this total wealth over a given time period, such as a year, which is an appropriate economic counterpart to the concept of sustainability, in that it includes not only natural resources but also (in principle at least) those other ingredients necessary to provide future generations an opportunity set that is at least as large as what is currently available to the present generations.

ANS is a special indicator to measure overconsumption and insufficient investment, which is derived from standard national accounting measures of gross national savings by making four adjustments. First, the estimates of the capital consumption of produced assets are deducted to obtain ANS. Second, current expenditures on education are added to net domestic savings as an appropriate value for investment in human capital (in standard national accounting these expenditures are treated as Consumption). Third, estimates of the depletion of various natural resources are deducted. Finally, global pollution damages from carbon dioxide emissions are deducted. It's a very simple addition and subtraction calculation, but why add and subtract? How much should each be added or subtracted? It's hard to conclude.

ANS means that Extended Wealth does not show as much as Gross National Savings. This indicator mainly makes reduction adjustment, but theoretically the calculation results are not

necessarily smaller. The key is how to price each adjustment item. Because the item has additions and subtractions. In the sense of numerical calculation, if the added item of education expenditure is relatively large and can offset the above 1, 3 and 4 items, adjustment results may increase instead. But in economic reality, the adjustment result usually does not increase. From an economic significance of view, ANS is generally less than GNS.

In indicator design, ANS emphasizes the concept of Economic Rent. The estimation of resource depletion is based on the calculation of Resource Rent. Economic Rent represents the additional return to a given factor of production, that is, the difference between the global price and the average unit extraction or harvest cost (including a normal return on capital).

Existing ANS calculations show that while the gap in levels between ABS and GS is mainly due to Capital Consumption and Human Capital accumulation whereas, according to the index, natural capital changes play only a relatively marginal role. This means that pricing of Resource Rent is low at calculation, which means that it is not enough to consider factors of sustainable development for global prices and average unit extraction or harvest cost, or the two have an offsetting effect.

Methodological shortcomings of ANS

The main methodological shortcomings of *ANS* are as follows.

First, the result of *ANS* depends on the number of adjustment items. The fourth adjustment is only limited to carbon dioxide emissions but does not include other important factors that lead to environmental degradation, such as underground water depletion, unsustainable fisheries, and soil degradation, and a fortiori biodiversity loss. One important reason is that the deduction amount is difficult to determine. How long can the list of pollutants be opened? Are there more identifiable pollutants under the current technical conditions? How can different countries coordinate a common and feasible list? How can economists coordinate with environmentalists? This is still a matter of measuring feasibility.

Second, the use of market prices to evaluate flows and stocks is warranted only in a context of perfect markets. If the economic reality is far from perfect competition, or the price does not exist at all, pricing issues may seriously affect the adjustment results. Externalities and uncertainties are paramount. If market price fluctuates sharply without smoothing, *ANS* will fluctuate sharply accordingly. Abandoning the market price, how to determine the accounting price? How to build a model? Less adjustment, no effect. With much adjustment, it is easy to negate the role of other capitals. What is important between human capital and physical capital? Added items or subtracted items, which is more important? What is the most important sub-item in added items and subtracted items? It is difficult to determine.

Third, it is difficult to determine its economic significance by computing *ANS* by country. One may feel uneasy facing with the message conveyed by *ANS* about resource exporting countries (e.g, oil). In these countries, from the *ANS* perspective, non-sustainability stems from an insufficient rate of reinvestment of income generated by the exploitation of natural resource: over-consumption of resource importing countries has not been reflected. Developed countries, which are generally less endowed with natural resources but richer in human and physical capital than developing countries, would then appear unduly sustainable. When computing *ANS* per country, the economic significance must be clearly explained. If the price of exhaustible resources in the international market fully reflects their scarcity, there would be no reason to make such a correction.

The importing country pays less for its imports than would be required; it will have a responsibility in global non-sustainability that is not captured by the money-value of its imports. Low

prices allow such countries to over-consume and to transfer the long-term costs of this over-consumption to the exporting countries.

However, some economists put forward that a part of the income from resource exporting countries should be devoted to the protection of resources and environment. Imagine that it would be excessive to require the poor countries with export resources to also consider compensating for renewable resources and protecting the environment. I am afraid it would be too much. The price of resources is so low that it is often difficult to maintain simple reproduction. They have no strength or energy to spare to protect resources and environment.

Fourth, *ANS* is still based on the concept of Net Product, which re-encounters the difficulty of measuring when GNP replaced Net National Product in the early years, that is, we cannot accurately separate the consumption value of Fixed Assets in each period. Now that *ANS* is advocated as the core indicator, but it is necessary to provide an explanation of whether the original measurement difficulty has been solved, and how has it been solved, otherwise the risk of rupture is implicit in the measurement logic. We can't pretend that circumstances will change as time goes by. Going back may be a sign of deep predicament.

The Adjusted GDP method is a main idea in the future well-being (sustainable development) measurement, but it is not the case in the current well-being (quality of life) measurement. This contrast has caused the author's doubts. If the method of correcting (adjusting) GDP cannot be used to measure the current well-being, how can it be used to measure the future well-being? Professor Diane Coyle believed that any kind of amendment to GDP is an attempt to transform it into something completely different from the original intention of the design (Coyle, 2014). This is a complete denial of "Adjusted GDPs".

4.3.4. Index of concerned resources

The Ecological Footprint (hereafter *EF*) measures how much of the regenerative capacity of the biosphere is used up by human activities (consumption). It does so by calculating the amount of biologically productive land and water area required to support a given population at its current level of consumption (Stiglitz et al., 2010).

On the supply side, bio-capacity is the productive of the biosphere and its ability to provide biological resources and services useful to humankind. A country's Footprint (demand side) is the total area required to produce the food, fiber and timber that it consumes, absorb the waste that it generates, and provide space for its infrastructure (built-up areas).

Economic analysis should focus on both demand and supply, so that the meaning of indicators can be understood more comprehensively. This is the most basic perspective of economics. No matter how detailed and deep the thinking is, the starting point is always here. If the train of thought is broken or blurred, we may as well continue it from now on.

The following is a further interpretation of the ecological footprint from seven aspects.

First, compare current consumption flow and its impact on environment with existing stocks. This is the general idea of sustainable development measurement, and the ecological footprint is measured according to this idea. It is also a wealth category measurement, but it only focuses on natural assets and is of course limited to physical wealth. The valuation rules of the ecological footprint are different from *ANS* and do not use market prices.

Second, this indicator shares with accounting approaches the idea of reducing heterogeneous elements to one common measurement unit (the global hectare, e.g., one hectare with productivity equal to the average productivity of the 11.2 billion bio-productive hectares on Earth). If a country's productivity is lower than the average level, it needs to occupy more area, resulting in an ecological deficit, otherwise, an ecological surplus will appear. It assumes that different forms of natural capital are substitutable and that different natural capital goods are additive in terms of land area.

Third, the EF is essentially a physical composite indicator, a special form of physical indicators. Some people regard it as a composite indicator for a certain reason, but it is essentially different from the dimensionless composite. The ecological footprint indicator is popular because it can express the pressure on the environment in an easy-to-understand unit. Easy to interpret is very important, and the indicators should have the ability to make people heart-wrenching.

Fourth, the EF indicator strongly opposes the weak sustainability assumptions. It does not consider the accumulation of savings and capital at all: any positive ecological surplus (bio-capacity that exceeds the EF) will not lead to an increase in the stock of certain natural capital, and hence the subsequent improvement in future production capacity. Saving and accumulating manufactured or Human Capital is not helpful in optimizing sustainability. On the other hand, the weakness of the EF is that it ignores the threat to sustainability resulting from the depletion of non-renewable resources, such as oil, and only consider waste assimilation factors (implied carbon dioxide emissions) rather than from an analysis based on the depletion dynamics.

Fifth, the major anti-trade bias is inherent in the EF methodology, for it does not consider that resources can be communicated, and environmental quality can be passed on. Ignoring this characteristic of the measured object, the result of indicator will be biased. The fact that countries with high population density (low bio-capacity), such as Netherlands, have ecological deficits, whilst countries with low population density (high bio-capacity), such as Finland, will have an ecological surplus. Trade is mutually beneficial, and the population can also move. Mobility is a sustainable way and can be measured. Ignoring these factors indicates that the measurement indicator is flawed.

Sixth, in recent years, research tend to avoid comparing a country's EF with its bio-capacity, instead, propose to divide all countries' EF by global bio-capacity. This means that EF is not a measure of a country's own sustainability but the country's share of responsibility for global unsustainability. Trade itself is a mechanism for sharing responsibilities and interests, which is intended to strive for achieving a win-win outcome. Therefore, the adjustment should be measurement of the share of responsibility.

Seventh, Carbon Footprint (*CF*) is one of the series of indicators of EF. It is generally acknowledged that the data quality is relatively high. The SSF suggests that a less encompassing but more rigorously defined footprint, such as *CF*, would seem to be better suited, since they are more clearly physical measures of stocks that do not rely on specific assumptions about productivity or an equivalence factor. As far as communication is concerned, such an indicator is just as capable of sending strong messages in terms of the over-utilization of the planet's capacity for absorption. The *CF* also has the interesting feature of being computable at any level of disaggregation. This makes it a powerful instrument for monitoring the behavior of individual actors. In generally, if the indicator covers a wide range, the certainty of its significance is poor, and its perceptibility is also poor. On the contrary, the boundary is narrow, and the result is relatively more assured. In this regard, the author has a special discussion in "Antinomy Between Gains and Losses in Economic Measurement". However, according to the functional requirements of the indicators, there is a tendency to expand

the measuring boundary. Does it satisfy functional requirements or focus on the quality of data results? The measurers are in a dilemma. The advantage of *CF* lies in the matching of function and quality. From the perspective of development, physical indicators often have evaluation advantages in some aspects of things being evaluated, and usually cannot give a general quantitative recognition.

4.4. Further discussion on the methodology of sustainable development measurement

The general methodological issues involved in the four indicators are discussed below.

4.4.1. What are we measuring exactly?

Since the Brundtland Report was published, the notion of sustainable development has widely accepted, but as an all-encompassing concept that absorbs every dimension of present and future economic, social and environmental well-being. It brought economic measurement into a dilemma.

The more important things are, the more measurements are derived from them. Different composite indicators convey different information. The problem is that there are many component indicators, many composite index and different information transmission, so there are choices and disputes about differences. It shows that many measurement standards are a major defect, which means a pattern of multiple data from multiple sources.

The SSF Report therefore emphasizes that it seems reasonable to separate the two notions of current well-being and of its sustainability. It concentrated on the sustainable component of sustainable development. We do want to end up with a limited number of indicators—a micro dashboard—and one that is specifically dedicated to the sustainability issue, based on a clear notion of what sustainability means.

The Economic Measurement Boundary Antinomy tells us that economic statisticians are often kidnapped by society. The government or the public need to know the quantitative performance of the things they care about, and the data results are totally negated if they are not in line with their expectations. Requirement is too much and too high, and in many cases, economic measurement has undertaken the impossible mission.

Such as Green GDP, which is also insufficient for assessing sustainability. The proximity that such a sustainability indicator would necessarily have with standard GDP could be a source of confusion. If there are two GDP indicators, which one should we use different economic contexts? What conclusion would we draw from the fact that a given country's Green GDP is $x\%$ or $y\%$ of its GDP defined in standard terms (Stiglitz et al., 2010)? On the contrary, Green GDP has caused more confusion and controversy in the measurement, which is one of the reasons why the trial results are not be published.

If we want to measure sustainability, what is required is a comparison between this concept of genuine production and current consumption. All it makes is the appropriate sustainability index more akin to a concept of net investment or disinvestment, and this is precisely the route that Extended Wealth or *ANS* exemplifies, but which is also followed by footprint indicators that are more specifically focused on the renewal or depletion of environmental assets. The argument goes as follows: the capacity of future generations to have standards of well-being at least equal to current generation depends upon whether we can pass them enough assets that matter for well-being. The logical starting point of the measurement is the definition of sustainable development mentioned when discussing the concept premise.

If we denote Extended Wealth index by “ W ”, which is used to quantify the stock of resource. Using Extended Wealth index to measure sustainability is equivalent to testing the direction of change in its global stock or its components, that is, computing its current rate of change, dW or dW_i . If the rate of change is negative, this means that consumption or well-being will adjust downward sooner or later. This is exactly what one should understand by non-sustainability. This understanding helps to eliminate misunderstandings about certain indicators. For example, ecological disasters reduce the resources available to create future wellbeing. This accounts for the fact that it deteriorates sustainability by decreasing the resources available for generating future wellbeing.

4.4.2. Summarizing sustainability in just one number: is it realistic?

To summarize sustainability in one number, we need to consider: what would be required to measure dW index in a satisfactory way. We must be more specific about several concepts: The goal—what is to be sustained? The mechanism—how do the various assets that will be passed on to future generations affect this measure of well-being? The share—What should the relative weight of each asset be?

That is more problematic, and it tends to clarify the opposition between the proponents of monetary indicators and physical indicators. In calculating dW , the weight problem highlights this contradiction. Is there actually a reasonable prospect of evaluating everything in money units? If all assets were traded in a perfect market by perfectly forward-looking agents that fully considers the welfare of future generations, one could argue that their current prices reflect the discount flow of their future contributions to future well-being. This is the ideal state of human harmony and an ideal premise of adopting monetary aggregates.

However, if many assets are not traded in the market, the current prices are unlikely to fully reflect this future-oriented dimension, due to the imperfection and uncertainty of the market, and short-sightedness. This implies that the true measure of sustainability requires a rate of change index, which assets are not valued at market prices, but rather at calculated accounting prices based on some objective physical or economic models of how future damage to the environment will affect well-being. At the same time, the current increase in human or material assets may help maintain and promote future welfare, which needs to be accurately assessed. In other words, to determine the accounting price, it is necessary to fully consider the positive effects of Human Capital and Material Capital and the negative effects of environment and resources. Such a perfect model must be constructed before credible results can be obtained.

There are two necessary conditions for evaluation according to Accounting Price: One is a full set of economic and physical forecasts on how the initial conditions determine the future joint path of economic, social and environmental variables. The other is to define a priori of how this path will behave in the future, that is, the knowledge of the social utility function. It is generally formalized as the discounted value of the sum of well-being in the future. Equipped with this tool, it should be possible to derive a sustainability index with the desired characteristics, that is, an ability to anticipate future declines in well-being below its current level.

To build this perfect model, we need to know the initial conditions, how the various variables interact, the length of time, the social utility function, and we need to formalize it to get the discount value of welfare in various periods in the future. Have we indeed met these two basic conditions?

Some simulations proposed in the technical report of the SSF Report illustrate certain aspects of this capacity. First, for those countries that are on unsustainable paths because of an insufficient output capital accumulation or renewal rate, this sustainability index is best suitable for issuing correct warnings. The calculated index should have this characteristic, which is only one of the characteristics that the index should have. Although environmental issues are of considerable importance, we cannot ignore other dimensions of sustainability. Second, such an indicator is inconsistent with a “strong” non-sustainability concept (i.e., problems arising from the depreciation of environmental assets that are essential to human well-being or even survival) only when it relies on fixed price levels for natural and unnatural assets (Stiglitz et al., 2010).

To understand the prediction of this kind of physical-economic model, please pay attention to the phrase used in the SSF Report, “If...If”. The premise is provided, under which premise this conclusion is reached. Therefore, the proponent has warned that it is the modeler’s responsibility to fail to meet the premise. Let’s see: if we were able to derive this index from a physical-economic model to predict the future interactions between the economy and the environment in a credible way, then this index would provide us with the correct warnings of non-sustainability, through strong increases in the relative accounting or imputed prices of these critical natural assets. The author believes that the problem lies precisely with so many “if”. This construction is still purely theoretical and utopian. It shows us at best the direction in which index builders should strive for.

However, it can also be used as a tool to highlight the many methodological obstacles to the construction of a comprehensive index. It proves that the author’s three questions about composite index are of great value. It is impossible to try to get the optimal solution, so we focus on the sub-optimal solution or even the satisfactory solution, which involves the issue of regulation. How to identify it? Who will judge? These issues will cause controversy.

In fact, the SSF Report has a very important judgment: a unidimensional view of sustainability certainly remains attainable. Because the weights of different items are chosen at random when constructing composite index, the consequences are seldom clearly stated. From an economic perspective, the author defines that the weight is the quantitative expression of preference. However, preferences are very mixed and difficult to express quantitatively, so the weight determination is the core issue of any economic quantitative analysis.

Expanding thinking, what is the general significance of the so-called unattainable conclusion of the SSF Report for composite index? Is it a denial of its feasibility? If so, is there any value in the existence of so many composite indicators in the world? At least we need to consider the value of its existence.

4.4.3. Technological uncertainty and synthetic methods

Measuring sustainability with a single rate of change index can work only under two strong assumptions. One is that the future eco-environmental developments can be perfectly predicted. The second is that there is perfect knowledge about how these developments are going to affect well-being. These two assumptions are quite different from the situation of the real world. The debate on the eco-environmental perspectives is dominated by ignorance and uncertainty about the future interaction between the two fields. There is a lack of consensus on the proper definition of the objective function. The SSF Report points out that uncertainty takes many forms, some of which are amenable to probability calculations. Probability tools can be used to remove uncertainty, but in some cases, they cannot. Probability logic is not omnipotent, and even mathematical logic is not omnipotent.

The impact of uncertainty on measurement affects not only the parameters of the models that one may try to use to predict eco-environmental interactions, but also the structure of the models themselves, the measurement of current stocks, and even the list of natural assets for which current and future stocks need to be taken into account. There will be different beliefs about future eco-environmental scenarios, which will lead to disputes on how to measure them, but there is no reason to believe that sustainability measurement should avoid the difficulties.

One solution to deal with uncertainty is to work based on scenarios or provide confidence intervals. Forecasters will do this when they intend to emphasize the uncertain nature of future trends. Another is the stress tests, that is, re-compute them under assumptions of external shocks on asset values. The external shocks include a sudden increase in the value of environmental assets, or a drastic reduction in the value of some other items--such as production capital or human capital. Such modes of presentation could be explored and eventually adopted.

The distinction between strong sustainability and weak sustainability is often unavoidable. The SSF Report points out that the aggregate indicators are inherently not suitable for measuring strong sustainability. Strong sustainability is a one-vote veto, so physical indicators are often used. Monetary aggregates indicator itself can also be used to calculate strong sustainability, but the standard is not so strong. The problem is that only by adopting extreme valuations for critical environmental assets can it be possible to respond to the challenge. Unfortunately, we are not well equipped to quantify precisely what these extreme values should be. In such cases, some projects cannot obtain a speculated monetary value, therefore, a separate independent physical accounting is inevitable.

The acceptability of indicators is also important. We accept the advantages of monetary aggregates because we are familiar with it therefore willing to use it. The magnitude of monetary aggregates is easy to understand and can be related to other currency values. Currency index has the advantage of using units that speak to everyone. In contrast, physical measurement indicators often require professional knowledge, such as atmospheric indicators, if the public does not understand them, it is difficult to use them to express sustainability. The EF index is better because the public knows the concept of area. In short, it is essential to find more suggestive ways to highlight the data.

4.4.4. Uncertainty also involves normative

The point is, what exactly are we going to sustain? There can be as many indicators of sustainability as there are normative definitions. We attempt to infer the definition of well-being by observing how people currently evaluate environmental factors relative to economic factors, using specific values or directly measuring the impact of environmental comfort on subjective well-being, forming an empirical approach.

Measuring sustainability with a single index number would confront us with severe normative questions. Normative and empirical analysis are inseparable. The two kinds of analysis influence each other. Can we solve this normative problem? Today's specific valuation and subjective degree of observation are based on our special ecological economic framework. Can the current framework be used to predict the valuation of future generations? They may face very different ecological economic structures.

Real economic life is full of examples of such normative issues. It could be argued that future generations may become very sensitive to the relative scarcity of some environmental goods, which we pay little attention to today because they are still relatively abundant. This requires that it be immediately

endowed with high value. The prices of many items are now too low, and future generations and even our generation want to raise their prices. If the price is high, we will not waste so much.

Another example of these normative issues is how sustainability indicators should add up to individual preferences. It depends on how to incorporate the distribution factor into the perspective of measures of current well-being. This would be in line with intra-generational differences that are often overlooked in Brundtland's definition of sustainability. We should pay attention to both intra-generational and inter-generational resource allocation in terms of sustainability, while focusing on intra-generational differences requires special attention to the bottom population.

If it is believed that the primary indicator of current well-being should be the disposable income of the bottom 80% of the population, rather than that of the total population, then sustainability indicators should be adjusted according to this objective function. Some people may not accept the statement of the bottom 80%. If the gap between classes is widened, compared with Warren Buffett and George Soros, 80% of the population is indeed at the bottom. Here, the core indicator is a structural analysis method that focuses on the key points. If we have the capital adequacy ratio indicator, and then design and calculate the core capital adequacy ratio indicator, we can better analyze the trend of capital adequacy ratio, because different types of capital have different transforming functions.

4.4.5. Another source of measurement complexity: the global dimension

The global dimension involves the real economic relationship between rich and poor countries. Advocates of the *ANS* argue that the problem of sustainable development generally concentrate in poor resource-exporting countries, even if AFC of resources is in developed countries. The logic of the argument is that, if the market functions properly, the pressure that developed countries exert on other countries' resources is already reflected in the prices that they pay for importing these resources. If import costs are removed, the *ANS* of developed countries remains positive, this means that they have made enough investment to compensate for the natural resources they consume. It is then the responsibility of resource-exporting countries to take a fair share of the income from exporting resources for reinvestment, if they also want to take on a sustainable path.

However, this implicit logic of the *ANS* holds true only under the assumption of efficient market. If the market is inefficient, and if natural resources are underpriced, then resource-importing countries benefit from an implicit subsidy while resource-exporting countries are essentially taxed compulsorily, resulting in price distortions and hence measurement distortions. The sustainability of resource-importing countries is overestimated, while that of resource-exporting countries is underestimated. This problem will be particularly serious when the market does not exist or presents strong externalities. The upstream countries of the industrial chain are the price makers, while the downstream countries are the price takers, with little room for choice or adjustment. From this, we can see that the motive why countries accelerate economic progress lies in striving for hierarchical advantages in the upper reaches of the industry chain.

The report assumes a two-country scenario, where both countries produce and consume with external effects on the stock of a natural resource that is global public good with free access, but its analysis is not very relevant to the opinions expressed in the report. Using extended wealth to evaluate changes and distinguish between the polluter and the polluted is certainly helpful for structural analysis, but there are also traps in interpreting its meaning. If factors such as importing for export and exporting for import are considered, further subdivision of the indicators is needed.

Environmental goods are public assets and are valued differently by each country. The message of this extended wealth concept is that the polluter is on a sustainable path, while the polluted is not. The message is clearly misleading, because it is difficult for the polluted country to reduce the negative impact of imported pollution and restore its sustainability, while it is only a technological change of the polluting country that could help restore the polluted country's sustainability.

It should be noted that economic analysis may not be able to exhaust all possibilities in advance. For example, an approach focusing on national sustainability may only be relevant to certain dimensions of sustainability, but not to others. Global warming is a typical example, as the expected consequences of climate change are unevenly distributed and may not be related to one country's carbon dioxide emissions.

It should also be noted that there are often other realistic settings that exceed the possibilities anticipated by the indicator designer. If we look at it from a global perspective, there must be some kind of "dirty" technology under the current technological level. Country *A*, which uses this technology, bears the responsibility of "doing evil" because of its low-end technology, while country *B*, which uses products of country *A*, is just in an advantageous position to evade responsibility.

Moreover, if a country is forced by the global division of labor, how should it allocate its national responsibility for sustainability, being both a polluted and a polluter? Resources can be transported, and the environment has no borders. There is a big difference between the global dimension and the national dimension, so we should pay more attention when measuring resources and environmental sustainability.

The SSF Report puts forward four suggestions on the measurement of sustainable development: establish a sub-dashboard, focus on reflecting the stock, focus on the economic aspects of sustainability in monetary aggregates, and use physical indicators to track and investigate separately. This is essentially a re-summarization of the previous discussion.

Particular attention should be paid to the last section of the report on sustainable development measurement, where two points are raised, which are essentially warnings to us:

First, social capital and institutional capital are not included in the dashboard, mainly because of the lack of consensus on how to measure them. The key lies in the direct consequence of this indicator caliber—total social capital can only stay in the theoretical concept, and there are major deficiencies in the concept of measurement or calculation. Usually, there are two main reasons for the lack of indicators or factors: Abandonment due to feasibility and low correlation, which are very easy to confuse. This is a trap often encountered in economic statistics.

Second, no limited set of figures can predict with certainty the sustainability or unsustainability of a highly complex system. The function of any figure is limited, and this warning makes us question the feasibility of sustainable development measurement.

In short, all four available methods are difficult to achieve. The dashboard method cannot give a single measurement result. The physical indicators are poorly integrated and can only be used in measurement of a single dimension. The composite index method is essentially unable to give an additive socio-economic explanation. The GDP adjustment method cannot overcome the difficulty of price measurement. It should be noted that the measurement of sustainable development seems to be well-established and is favored by all aspects of society. In fact, it is just hovering at the starting point of exploration. There is a long way to go to measure progress.

5. Conclusions

OECD has played a key role in promoting the implementation recommendations of the SSF Report worldwide. In 2011, OECD launched the “Better Life Initiative” to promote the “Beyond GDP” agenda. In 2012, OECD launched the “New Approaches to Economic Challenges Initiative” to reflect and improve the “Economic GPS”, which is the economic fluctuation detection system. In the same year, OECD also launched the “Inclusive Growth Project” and established the “OECD Framework for Policy Action on Inclusive Growth”. Based on the above background, OECD has established the “High-Level Group on the Measurement of Economic Performance and Social Progress” (HLEG) to study economic measurement.

The SSF Report has a latest and updated version issued in 2018, that is the SFD Report, divided into two volumes: One is the chairman’s volume, entitled “Beyond GDP—Measuring What Counts for Economic and Social Performance”; The other is the expert’s volume, entitled “For Good Measure: Advancing Research on Well-being Metrics Beyond GDP”. The paper believes that after the 2008 global economic crisis, the SFD Report is closer to economic reality than the SSF Report. In contrast, the SSF Report is more in transformation stage of measurement concept, which is divided into three parts: classical GDP issues, quality of life, sustainable development and environment, there is not much deviation in the three parts. The SFD Report draws lessons from the global economic crisis and highlights the measurement of downward trend of the economy. Its focus is on income distribution and inequality measurement, which is further divided into four aspects. The SFD Report also highlight the measurement of economic insecurity and trust.

It can be seen from content proportion that the focus of the SFD Report has changed greatly. This confirms the author’s view that sustainable development is a luxury, valuable but expensive. After economic downturn, its importance is relatively reduced. It is particularly important for developing countries to realize this, especially for emerging countries. Otherwise, in international competition of environmental protection, poor countries will pay a price beyond their economic strength, which may greatly slow down their process of getting rid of poverty.

Acknowledgement

This paper is a phased achievement of the Major Program of National Philosophy and Social Science Foundation of China (Grant No.18ZDA123), “Innovation Team of Philosophy and Social Sciences in Henan Colleges and Universities ”(2017-CXTD-07), and “Major Projects in Basic Research of Philosophy and Social Sciences in Henan Colleges and Universities” (2019-JCZD-002).

Conflicts of interest

The authors declare no conflict of interest.

References

- Coyle D (2014) *GDP: A Brief but Affectionate History*, Princeton: Princeton University Press.
- Clerc M, Gaini M, Blanchet D (2011) *Recommendations of the Stiglitz-Sen-Fitoussi Report: A few illustrations*, The French Economy.

- Deaton A (1997) *The Analysis of household Surveys: A Micro-econometric Approach to Development*, Baltimore: Johns Hopkins University Press.
- Deaton A (2013) *The Great Escape: health, wealth and the origins of inequality*, Princeton, NJ: Princeton University Press.
- Harari YN (2017) *Homo Deus: A Brief History of Tomorrow*, Harper Collins Publishers.
- Lequiller F, Blades D (2014) *Understanding National Account*, 2nd ed, OECD Publishing.
- Mankiw NG (2015) *Principles Of Economics—Macroeconomics (Chinese Version)*, Peking University Press.
- Masood E (2016) *The Great Invention: The Story of GDP and the Making and Unmaking of the Modern World*, Pegasus Books.
- Qiu D, Song XG (1999) Theory of Levels in Sustainable Development. *Stat Res* 2: 14–26.
- Qiu D, Chen MG (2007) There Should Be Less Self-accusation on Chinese Resource Consumption: Pondering on Resource Consumption Stratification Hypothesis. *Eco Res* 2: 66–71+81.
- Qiu D (2012) The Boundary Antinomy of Macro-measurement and It's significance. *Stat Res* 8: 83–90.
- Qiu D (2018) *Reflections on The Subject of Economic Statistics*, American Academic Press.
- Qiu D (2019a) *Logical Mining of Economic Measurement: Difficulties and Principles*, American Academic Press.
- Qiu D (2019b) Debate on Function of Inventing GDP of the 20th Century, *Logical Mining of Economic Measurement: Difficulties and Principles*, American Academic Press.
- Qiu D (2019c) Analysis of Several Measuring Dilemmas Implicated in GDP Statistics, *Logical Mining of Economic Measurement: Difficulties and Principles*, American Academic Press.
- Stiglitz JE, Sen A, Fitoussi JP (2009a) *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Available from: www.stiglitz-sen-fitoussi.fr.
- Stiglitz JE, Sen A, Fitoussi JP (2009b) *The Measurement of Economic Performance and Social Progress Revisited: Reflections and Overview*. Available from: www.stiglitz-sen-fitoussi.fr.
- Stiglitz JE, Sen A, Fitoussi JP (2010) *Mis-measuring Our Lives: Why GDP Doesn't Add Up*, The New Press.
- Ward M (2004) *Quantifying the World-UN Ideas and Statistics*, Indiana University Press.
- Wilkinson N, Klaes M (2012) *An Introduction to Behavioral Economics*, 2nd ed, Palgrave Macmillan.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)