Quantitative Finance and Economics

QFE, 5(4): 591–603.
DOI: 10.3934/QFE.2021026
Received: 03 June 2021
Accepted: 24 August 2021
Published: 27 August 2021

http://www.aimspress.com/journal/QFE

Research article

Paradox in deviation measure and trap in method improvement—take international comparison as an example

Dong Qiu[1] 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: Due to the problem of "true value agnostic" in the measurement of the real world, people believe that the existing methods can be closer to the true value by improving them. Therefore, they are willing to excessively affirm the more advanced method and deny the relatively "traditional" method. Taking the exchange rate method and purchasing power parity method commonly used in international economic comparison as examples, this paper generalizes the problems revealed by the exchange-rate-deviation index and concludes that there are at least three paradoxes in the deviation measurement of different methods. These paradoxes are the paradox of behavior significance, the paradox of comparative object, and the paradox of measurement result. The reason is that there is a cautionary trap in the improvement or innovation of measurement methods in reality. Sometimes the improved method is not necessarily better than the unimproved method. People tend to prefer advanced technology and methodology, but the problem of statistical input and related statistical benefits need to be considered in practical measurement. In fact, these basic problems still exist in some of the methods of economic statistics that we regard as common sense. When learning or introducing new methods, scholars do not absolutize the existing methods and conclusions. They should pay attention to critical experience, avoid the trap of improvement methods, and seek real improvement or innovation.

Keywords: deviation measure; paradox; method improvement; trap

JEL Codes: G15, F36, C40

1. Problem presentation

In the measurement of the real world, people always encounter the problem of "true value agnostic". People always hope to be close to the true value and have more accurate methods. And we constantly innovate to improve the original methods or to discover or invent new methods. It is a long process of pursuing the truth.

When people have different measurement methods, they will naturally compare the existing methods. In the comparison process, people usually have a tendency: they are willing to excessively affirm the more advanced method in the existing methods and deny the relatively "traditional" method. When we compare the two methods, we often regard the results of the advanced methods as the benchmark. Sometimes, when we calculate the gap between the results of the "traditional" methods and the results of the advanced methods, we regard this gap as the deviation of the "traditional" methods from the "true value" of things.

There are actually pitfalls in the improvement of measurement methods. And it is difficult to detect the paradox of the measurement deviation of different methods. Let's start with the method of international economic comparison.

We know that the exchange rate method and the international comparison project (ICP) method are mainly used in international economic comparison. These two methods both deal with the conversion of currency purchasing power of economic indicators in different countries (Qiu, 1996). The difference lies in different methods, and the degree of conversion may be different. But the emergence of the ICP method stems from the doubt about the results of the exchange rate method. When it was found that the results of the exchange rate method might be different from other evaluations, people took an extremely negative attitude towards the exchange rate method. And they decided to start all over again, spending a great deal of human and financial resources to develop a new international economic comparison method (i.e., ICP method). With the deepening of ICP research, the exchange rate method has gradually lost its dominant position in international economic comparison. Moreover, the measurement of the exchange rate method now exists only as a rough reference value. The World Bank's *World Development Report* has published the GNI data of more than 100 countries in the world calculated by the exchange rate method and ICP method (World Bank, 2004). We can see that there are obvious differences between the results of the two methods. The difference between the results of the two methods is usually measured by one indicator in the field of economic statistics, which is the exchange-rate-deviation index or exchange-rate-deviation coefficient.

How to view this indicator? How to look at such a big gap between the two methods? Some scholars believe that the deviation between exchange rate and purchasing power parity is inevitable and reasonable from both theoretical and empirical perspectives. By comparing the differences between the exchange rate method and the purchasing power parity method through the ICP report data, it is calculated that the exchange-rate-deviation of developed countries is smaller than that of developing countries, and the exchange rate deviation of developing countries is gradually narrowing over time (Yu, 1995; Ma et al., 2016). There are many reasons for the deviation between the exchange rate method and the ICP method, such as the defects and deficiencies of the theoretical system of purchasing power parity (Chen, 1993), the disunity of the value scale in the ICP summary calculation formula (Peihua, 2007), and the improvement of summary method (Wei, 2018). As well as the influence of various national institutions and economic, social and cultural factors, the ICP program has many challenges in practice (Chen, 2020; Ferrari et al., 2005). Therefore, the quality of survey data, comparison methods and results also have certain limitations. Caution should be taken in the application of the comparison results, and the world bank ICP results should be viewed correctly (Yu, 2015; Wang, 1993). It has been suggested that combining the results of the two methods can provide a better evaluation of the economic quality of a country (Ma et al., 2016).

Other scholars pointed out that the consideration of this issue should start with the shortcomings of the exchange rate method, because the purchasing power parity method (i.e., ICP method mentioned in this paper, considering that the measurement of exchange rate method is based on the theory of purchasing power parity, so the method adopted in this paper is not called this—note to the author) is an improvement on the exchange rate method. The difference between the two methods must mainly come from the shortcomings of the exchange rate method. For example, the exchange rate fluctuates greatly and frequently in the short term (Bo et al., 2001), the exchange rate system is diverse and complex, the factors that determine the exchange rate fluctuation gradually increase (Wang,1995), and the scope of non-tradable goods is not involved (Yu, 2003). The exchange-rate-deviation index has the econometric significance of measuring the error of exchange rate estimation.

From the general logic, it is natural to make the above judgment. However, we should note that international economic comparison is also a "true value agnostic" question. The exchange rate method itself does have various drawbacks, such as violent fluctuations and government interference, which may lead to results contrary to the real economic relationship. But some criticisms often apply to the ICP method. And we have not been strictly proved that the ICP result itself is the absolute standard to measure the accuracy of the international economic comparison. It's probably just a little closer to the "true value". The ICP result may also have errors in the "true value" of international economic comparison. In linear comparison, the bias may be positive or negative.

Therefore, the so-called "exchange-rate-deviation index" is unscientific. A premise of this index design is to take the ICP results as the benchmark to judge whether the international economy is more accurate or not. Regard these results as the "true value" and measure the bias caused by the exchange rate comparison. In fact, the difference between the exchange rate method and ICP method results is not equal to the error of the exchange rate method itself. Of course, the results of the exchange rate method will be biased. Still, ICP results may also be biased in international economic comparison, and the direction of bias may be different from that of the exchange rate method. Thus, the error of the exchange rate method in international economic comparison may be less than the distance between it and ICP results. For example, if the true value of the two countries' purchasing power parity is 8:1, the exchange rate method result is 4:1, and the ICP result is 10:1. The error of the exchange rate method result should be only four units, not six units. From this point of view, we should pay attention to avoid being misled when applying the so-called "exchange-rate-deviation index" data.

However, in various works of international economic comparison, there is a tendency to deny the exchange rate method. In contrast, the ICP method is used in the world under the aura of science. So far, ICP's primary literature has not elaborated systematically and deeply on the comparative mechanism of the ICP method (Qiu, 2020). Most of the so-called research is just its application, or at most make some repair on the aggregation method. Many economists do not study the measurement process of the ICP method in detail, so they believe in the results of the ICP method. They use it on many occasions to analyze the actual economic problems of different countries and put forward their countermeasures and suggestions. Their authoritative assurance led to the general public's assurance of the ICP results. It seems that the conclusion of ICP has become unquestionable and indisputable.

The question is, are the results of ICP as good as we believe they are? Are the results of the exchange rate method as bad as we believe? Is exchange-rate-deviation index the deviation from the true value of the purchasing power ratio of money caused by the exchange rate method? In 1996, the author wrote an article to raise doubts. Today, twenty-five years later, the author further analyzes this problem from the perspective of the deviation measurement paradox and method improvement trap.

2. Deviation measurement paradox

By generalizing the problems revealed by exchange-rate-deviation index, we believe that there are at least three paradoxes in deviation measurement.

Paradox 1: Paradox of behavior significance

From the definition of indicator, strictly speaking, the true value of an indicator is agnostic. It is consistent with Heisenberg's Uncertainty Principle (Eastwood, 2017), or it can be verified.

It is agnostic in theory but needs to seek knowledge in practice. The vast contrast between the two determines that we can only try our best to approach, that is, use the measured value of the indicator to approach the true value of the indicator as much as possible. Different approximation methods produce different measurement values, which leads to the accuracy comparison or deviation calculation of different measurement methods.

By definition, the deviation should be the difference between the actual value (measured value) and the true value. Here, the true value becomes the means to calculate the deviation. According to this definition, if you don't know the true value of the indicator, you can't know the deviation of the indicator, and the deviation is as unknowable as the true value.

However, if the true value is known, there is no need to calculate the deviation. In this case, the design of this concept loses its behavioral significance. It is a deviation uncertainty theorem in the sense of concept or behavior or a deviation paradox in the behavioral significance.

Paradox 2: Paradox of comparative object

From the paradox of behavioral significance of deviation measurement, we can see that only when people take the result of a certain approximation method as the true value, the result of other methods will form the so-called deviation. The calculation of any deviation can only be the comparison of two measures (A and B).

Similarly, since the true value is agnostic, we can't determine that one measured value must be better than another measured value in advance, but the original intention of the design and calculation of deviation measurement is: one of the two compared values should be the benchmark value, or require a priori determination of a benchmark, or at least require one of the two measured values to be closer to the true value.

Therefore, another paradox arises. If the benchmark value (assumed to be B) exists and is known (or knowable)—in other words, B is known to be better than A, then another measured value A is unnecessary to measure, or the necessity of measured value A is insufficient from a certain point of view. It is the deviation uncertainty theorem from the comparative object or a deviation paradox in the comparison object.

Paradox 3: Paradox of measurement results

Only in the linear state of plane space we can find the deviation uncertainty theorem from the measurement results or the third paradox of deviation measurement—the paradox of measurement result.

Let's consider three values (true value T, measured value A and measured value B). Measured value A is measured by the original method, and measured value B is calculated by the improved method B. It is generally believed that measure method B is better than measure method A, so measured value B should be closer to the true value than measured value A, right?

If we analyze this problem from different angles or factors, we may find different relationships among them.

First, they may overlap in very accidental situations, or overlap in three, or two (A and T, B and T, A and B).

Second, in the linear state of plane space, the measured value A and B can approach the true value T from different directions or the same direction.

Third, there may be different relations among the magnitude of true value T, measured value A and measured value B in the linear state of plane space. we cannot know the magnitude of true value, but we can infer the relationship between true value and measured value.

Table 1. Various relationships among true value T, measured value A and measured value B.

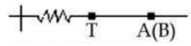
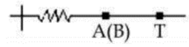
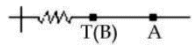
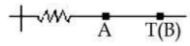
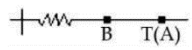
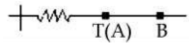
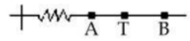
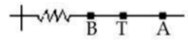
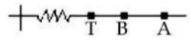
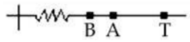
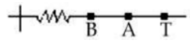
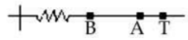
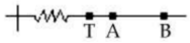
Number	Overlap Degree	Approaching Direction	Value Amount	Result	Graph
(1)	Complete overlap			B = A = T	
(2)	A = B		A = B > T	A − T = B − T	
(3)	A = B		A = B < T	T − A = T − B	
(4)	B = T		A > T = B	B − A = T − A	
(5)	B = T		T = B > A	B − A = T − A	
(6)	A = T		A = T > B	B − A = B − T	
(7)	A = T		B > A = T	B − A = B − T	
(8)	No overlap	Different	B > T, B > A, T > A	B − A > T − A	
(9)	No overlap	Different	A > T, A > B, T > B	|B − A| > |T − A|	
(10)	No overlap	Same	T > A, T > B, B > A	B − A < T − A	
(11)	No overlap	Same	A > B, A > T, B > T	|B − A| < |T − A|	
(12)	No overlap	Same	T > A, T > B, A > B	T − A > A − B	
(13)	No overlap	Same	T > A, T > B, A > B	T − A = A − B	
(14)	No overlap	Same	T > A, T > B, A > B	T − A < A − B	
(15)	No overlap	Same	B > A, B > T, A > T	B − A > A − T	
(16)	No overlap	Same	B > A, B > T, A > T	B − A = A − T	
(17)	No overlap	Same	B > A, B > T, A > T	B − A < A − T	

Table 1 is a summary of the relationships among the three values after considering the three factors.

Table 1 lists 17 possible cases of the relationship between the three values, from which we can get the following understanding:

In case (1), the three values (T, A and B) overlap completely. In case (2) and case (3), two measured values (A and B) overlap, and there is no deviation measure problem. Of course, this case is very rare.

In case (4) and case (5), the measured value B overlaps with the true value T. The original understanding of deviation and the recognition of measure B method are correct, that is, $|T - A| = |B - A|$. But we should see that: first of all, this is also a very rare case. Secondly, we can't know whether or when these two cases happen.

In case (6) and case (7), the measured value A overlaps with the true value T, but the more accurate measured value B is far away from the true value T, that is, $|B - A| = |B - T|$. At this time, the original conclusion is not tenable. Of course, this is rare, and its existence and timing are unknown.

In case (8) and case (9), two measured values A and B approach true value T from different directions, and the calculated deviation is greater than the true deviation, that is, $|B - A| > |T - A|$. We get a deviation measured value that has its deviation. Our original understanding of deviation is wrong.

From case (10) to case (17), two measured values (A and B) approach true value T from the same direction. In this case, the relationship between the calculated and true deviation is more complex, and three kinds of results may be produced.

In case (10), case (11), case (12) and case (17), the calculated deviation will be less than the real deviation, namely, $|B - A| < |T - A|$, and the calculated deviation itself has calculation deviation.

In case (14) and case (15), the calculated deviation will be greater than the true deviation, that is, $|B - A| > |T - A|$, and the calculation deviation will also exist.

In case (13) and case (16), the calculated deviation is exactly equal to the real deviation in quantity, that is, $|B - A| = |T - A|$, which is also quite rare.

When two measured values (A and B) approach true value T from the same direction, the case (10) and case (11) are consistent with the original understanding, that is, measure B is an improvement of measure A, which is closer to true value T than A. But in the six cases from the case (12) to case (17), the conclusion is the opposite. Measured value B is not an improvement of measured value A, but rather, measured value A is closer to the true value T.

To sum up the (14) cases (from case (4) to case (17)) with deviation measure, we can think from two perspectives: one is whether the direction of the original conclusion is consistent with the reality, the other is the magnitude of the deviation between the magnitude of the original conclusion and the reality, and then get the overall understanding, as shown in Table2.

Table 2. The consistency comparison of approaching direction, the measured value and the real value.

Value / Direction	Consistent		Inconsistent		
Consistent	case (4)	case (5)	case (10)	case (11)	The calculated deviation is less than the real deviation
Inconsistent	case (6)	case (7)	case (12)	case (17)	
	case (13)	case (16)	case (8)	case (9)	The calculated deviation is greater than the real deviation
			case (14)	case (15)	

It can be seen from Table2 that the deviation measure is consistent with the original understanding in the result direction and magnitude, only in case (4) and case (5), but as mentioned above, it is quite rare. Six cases are consistent in magnitude, which is also quite rare, and four cases are consistent in direction. On the contrary, there are eight cases of inconsistency in magnitude, including four cases where the calculated deviation is less than the real deviation and four cases where the calculated deviation is greater than the real deviation. There are ten cases of inconsistency in direction and six cases of inconsistency in both magnitude and direction.

It can be seen that the true value of deviation is to be calculated, but the calculated value is usually the value with deviation, and the result is contrary to the purpose. Paradox 3 can also be called the deviation uncertainty principle in the significance of measuring the result.

Above, we have pointed out various problems that may appear in deviation measurement. Of course, this is probably only one part of them, but only these problems are enough to make us ask questions and think more deeply: why should the advanced method B be inferior to the original method A? The reason why method B is proposed is that the result of the method A is not ideal and needs to be improved. How can it be better to improve but not to improve? We think that the reason for this possibility lies in: there are traps of method improvement or innovation in the real world.

3. Trap in method improvement

3.1. Traps of method improvement

Compared with simple methods, technologically advanced methods often need more presuppositions to ensure the accuracy of the results. Generally, the more advanced the technology is, the more presuppositions are required. When innovators improve or innovate methods, they are very concerned about the necessary preconditions and usually consider them carefully. The beginners of the new method also pay more attention to its preconditions to accept the new method from the heart. However, when the new method is accepted more and more, its wider range of users tends to ignore its preconditions.

However, in reality, data acquisition can't meet these assumptions. To achieve or complete the calculation, we must relax these assumptions. Judging from the PPP calculation mechanism, the quality of ICP implies three basic requirements: pure price ratio, equivalent ratio and mean representativeness (Qiu, 2018). In other words, to believe the ICP data results, we need to assume that all three requirements of basic heading PPP are met. However, there is considerable uncertainty in price measure, and even the "uncertainty theorem of price" can be asserted. Therefore, the foundation of ICP based on price adjustment is not stable.

This kind of relaxation will damage the accuracy of the results to a certain extent, which may lead to the trap of method improvement: because improvement or innovation can improve the accuracy or accuracy of a method $\Delta + B$. But in practice, relaxing the assumptions and preconditions of the innovative method will lose accuracy or accuracy $\Delta - B$. If the latter exceeds the former, that is $\Delta + B < \Delta - B$, then the improvement or innovation of method B is not as good as the original method A. At this time, the result of improvement or innovation is not as good as the original result without improvement or innovation. In other words, further improvement or innovation is needed to get more accurate results.

From the perspective of space, the more advanced the technology is, the narrower the effective space scope of its application will be. The mismatch between the effective space (small space) of the advanced technology and the real space (large space) leads to the failure of the advanced technology and method outside its effective space. People often ignore the difference between the two spaces. The effective technical methods in a certain space are applied in the failure space, and the application results are blindly recognized, which leads to the trap of method improvement or innovation.

In real life, there are also examples of technological innovation traps: it is difficult for precision VCD machines to play discs with rough quality, but VCD machines with slightly worse quality can "ultra-strong error correction", even if the discs are a little rough, they can play them. Precision water pipes are easy to cause problems when they are used for water with more impurities. On the contrary, not-so-good water pipes can better adapt to the erosion of low-quality water; This is the Enlightenment of real life, revealing the truth of this kind of improvement or innovation trap.

3.2. The influence of "method omnipotence theory" on measurement

People tend to worship technology and are more willing to believe and adopt methods with higher technology content. When people evaluate a certain measure, they often infer the reliability of the result according to the excellence of the method itself. The usual inference is: the method is exquisite, and the result is accurate. Whether the method is exquisite is only a necessary condition to obtain accurate results, but not a sufficient condition. If the necessary premise is not satisfied, it is likely to use a particularly exquisite method to get a consequence full of mistakes. When people compare different methods and their results, they are often easy or willing to forget the presupposition implied in improvement or innovation method B, or do not realize the existence of technological innovation trap. What's more, it's more terrifying than using the rough method to get the result because the beauty of the method often makes it easier to be sure of its possible wrong result. That is to say, people's tendency toward technology worship aggravates the harm of the method improvement trap.

When evaluating the international economic comparison results, too much negation of the exchange rate method and too much affirmation of the ICP method are misled by "method omnipotence theory". From the technological content of the method, the exchange rate method is relatively rough, while ICP is more refined and has the advantage of the form. Due to people's technical preferences, methods with high technological content are often more likely to be accepted. In the face of complex comparative objects, the ingenious method of ICP is farfetched and requires more preconditions (Qiu, 2020).

What's more, the use of computers brings convenience to people, but it also makes people ignore the supervision of the method application process and do not go into the in-depth investigation. What assumptions and preconditions are relaxed? How relaxed is it? How much influence does this relaxation have on the calculation results? When people use computer software, they actually regard the calculation process as a black box. They pay more attention to whether the method is good and how to analyze the results but do not have more in-depth knowledge of the calculation process. The famous judgment that the ICP method must be better than the exchange rate method has not been proved by academic research. And its comparison mechanism needs to be further explored in order to improve the standard of international economic comparison.

One of the consequences of relying too much on technology and methods is the possibility of a "comparative paradox": in reality, sometimes people are comparing incomparable things. As we all know, the premise of comparison is that the items to be compared must be comparable. But in practice, people may have to compare items that are not comparable or compare things that are not. In fact, ICP is based on the assumption that people are more intelligent than markets and can accurately identify the relationship of international purchasing power comparison. Abandoning the observations of market exchange rate instead of estimating the purchasing power price ratio in the market is essentially a matter of comparing the incomparable. The core difficulty encountered by ICP is the contradiction between "representativeness" and "comparability", which is the cost of comparing incomparable things (Qiu, 2020).

In the article "Some Thinking of International Economic Comparison", the author points out the danger that the comparable economic factors among countries will gradually decrease and even lose. In the case of purchasing power parity, it is impossible for ICP to compare all commodities, but can select only some representative specifications. In this regard, the basic premise of applying the ICP method is that the selection of representative specifications should follow the criterion of "centralized selection" and the "commonality" (Qiu, 1996). In order to ensure the representativeness of the sample, the representative specifications are required to be the commodity with the largest expenditure in the category. Commonality requires that the selected specifications be used in common with the reference country.

Without representation, it is impossible to measure the real income level of the reference country. Without commonality, the comparison is meaningless. However, these two criteria are often difficult to achieve at the same time.

Due to the level of economic development, natural conditions, resources, management system, living customs, and many other aspects, the commodities used by countries worldwide have certain differences in their per capita total amount and composition (Xu et al., 2021). Their differences vary in level and size, which form a sequence. In this sequence, the comparison between relatively similar countries can meet the requirements of representativeness and commonality, and the result is relatively ideal. The results will be biased by comparing countries farther apart in the sequence. The farther the difference, the worse the comparability. Finally, it is actually to put the incomparable things together and forcibly compare them.

3.3. On the choice of measurement methods from the perspective of statistical benefits

Advocating technology and methodology may also be the only criterion for judging and choosing the advanced technology and methods. But in practical measurement, there are still some important factors that must be included in the field of vision. It is the problem of statistical input and related statistical benefits.

First, it is the limitation of accuracy of comparison results of funds. The principles of economics tell us that there is no free lunch in the world, and neither is statistical measurement. Take the international economic comparison as an example. If the funds are insufficient, the number of countries participating in the ICP comparison is small, or the information provided is incomplete and inaccurate. The measures that should be taken to reduce errors are challenging to implement. In this way, ICP is in a dilemma: if we want to minimize the comparison error and improve the representativeness, it will inevitably lead to the complexity and high cost of the comparison process. The more complex the process is, the more expensive it is, the less feasible it is to implement the comparison. It is an ironic reality that only developed countries such as Western Europe and North America are able to guarantee the routine implementation of ICP. Only when ICP is not much needed (for developed countries, the gap between ICP results and exchange rate method results is small) will they have the economic strength to carry out ICP.

Second, the timeliness of measurement results. From the perspective of timeliness of ICP results, standard ICP usually takes place only after several years, and the duration of each ICP is relatively long, so the data release is seriously delayed. The exchange rate method is easy to operate, sensitive to the market, and has good data timeliness (Xu et al., 2017). Therefore, only using the ICP method for international economic comparison needs to solve the timeliness of the problem in order to be able to perform the economic development of each country timely.

Third, the controllability of the measurement process. In international economic comparison, there may be artificial adjustment in the collection of price data. ICP requires the reference countries to report the average price data, and each country has great flexibility in reporting the data. ICP is unable to avoid government intervention in price. Governments provide the basic price information required for the project. Meanwhile, there is great room for adjustment in the selection of sample cities and average methods. ICP is also in a dilemma: if the results are used for policy purposes, countries will have a tendency to "revise" the data; However, if ICP is only used for research purposes, countries have little interest and investment in it.

From the perspective of statistical benefit, ICP is not necessarily superior to the exchange rate method. Although the exchange rate method is relatively simple and the result seems unconvincing, it needs little investment; In contrast, the funding, manpower, and time required for ICP are amazing. The investment will be more considerable if we want to get more accurate results and meet the premise and nature of the comparison.

In this way, there is a balance between the accuracy of the result and the size of the input. Is it true that the higher the accuracy, the better? Can this be done in practice? Usually, the cost constraint has a great impact on the implementation of statistical measures, which is a hard constraint, and it is often necessary to seek the accuracy of measurement results under this constraint. How to determine the standard of method selection, accuracy, timeliness or economy? Different uses should have different selection tendencies. Considering only one factor will cause problems in practice.

4. Responsibility and value of the catcher in the rye

4.1. Back to the original

What this paper discusses is not the technical level of the economic statistics method but a basic problem. It seems inappropriate to look back at this the original when the ICP method has become popular. However, how to answer this basic problem is related to the future trend of the ICP method and exchange rate method. It is also related to the basic judgment of the analysis of international economic relations. In fact, it can't be bypassed. This is the invisible debt we owe. We always have to pay it back. This is the duty of our scientists. It must be pointed out that back to the original is not the first one in this paper. It is actually a trend worthy of our attention in international scientific research.

It should be noted that although this article takes the international economic comparison as an example to start the discussion, it is not only in the international economic comparison that there are such basic problems that need further consideration. In fact, these basic problems still exist in some of the methods of economic statistics that we have come to regard as common sense (Qiu, 2004). For example, in the measurement of business cycle fluctuation, when discussing the traditional methods, our textbooks all take the determination of the benchmark cycle as a very mature method. Many scholars also prefer to use modern measurement methods, that this kind of technological progress can better measure the actual business cycle, but both have benchmark problems.

4.2. The benchmark problem in the measurement of the business cycle

In the traditional business cycle measurement method, a very basic problem is the determination of the benchmark cycle, which involves two aspects: one is to help select the economic indicators according to the benchmark cycle, which can be used as cycle measurement indicators; the other is to determine the category of cycle indicators according to the benchmark cycle, whether an indicator is a leading indicator, synchronous indicator or lagging indicator of cycle measurement. In the determination of cycle indicator, there are six selection principles, two of which are related to the benchmark cycle. One is the time of the cycle, and the other is the consistency of cycle direction (). Wesley C. Mitchell said, “business history repeats itself, but always with a difference. Every business cycle, strictly speaking, is a unique series of events and has a unique explanation.” (Mitchell, 1913). If the benchmark is not absolute, will there be any bias in the selection of periodic indicators? And what is the confidence level of the three categories of the so-called “lead, synchronization and lag”?

Some people will say that the traditional method of measuring cycle fluctuation is not used, and its modern method does not have the so-called benchmark cycle problem. In fact, there are also benchmark problems in modern methods. We know that the modern method of measuring cycle fluctuation is to eliminate the influence of long-term trend, seasonal change, and irregular change from an economic output sequence (usually GDP) to get the periodic fluctuation of the economy. The use of modern methods actually implies two premises: (1) taking GDP as the benchmark of the whole economy, but GDP is constantly adjusting the measurement method with the change of economic reality; (2) there are only four economic change ways: long-term trend, periodic change, seasonal change and irregular change, because the measurement of periodic change is mainly based on the residual method. Only when the measurement of the other three changes is accurate will the business cycle measurement be accurate. The original here is: is there only four types of changes in economic operation? If there are other types of economic changes, then using random methods to measure irregular changes will have a systematic deviation (Lv, 2006). It can be seen that if these two premises cannot be fully met, our measurement of the business cycle is actually based on the castle in the air. The so-called modern method is not more reassuring than the traditional method in this point, although it belongs to the parameter method.

4.3. Critical experience should be paid attention to in learning new methods

A man named Hong Du said: “the aesthetic experience of man is unreliable, A natural need for fresh stimulation as the motive force, always go on the path of the pursuit of the tide. The true and reliable is the spirit of man’s negation and exploration” (“Chinese National Geography”, 2005 supplement. p.21). I think it’s true, and it’s probably the same problem with aesthetics in scientific research.

For scholars who are weak in disciplinary discourse power (mainly in developing countries), the critic’s self-consciousness is not intense. The consciousness of students’ identity is intense, which is easier to be conquered by the scientific aesthetic experience.

In particular, Chinese culture has a compatible tradition. The humility and inclusiveness inherent in the Chinese blood further aggravate this possibility. Because of this, Chinese scholars should remind themselves of more emphasis on critical experience. Or, talk in an extreme way, we should pay attention to the experience of ugliness in introducing modern methods.

4.4. The value of scholars

What is the value of scholars? It lies in their unique vision. They can be a watchman in science, find defects in previous achievements, see whether the methods and data support each other, point out the relativity of methods and conclusions, and warn society, especially the users of methods and conclusions, to pay attention to problems and defects. So, the existing methods and conclusions should not be absolutized. Scholars cannot drift with the tide. They cannot confuse themselves with ordinary citizens, nor can they become the advocates of false academic bubbles.

What's more, this paper puts forward the theory of deviation uncertainty (deviation measurement paradox), which is not to give up the measure of deviation, but to better approach the true value. This paper points out that the trap of method improvement or innovation is not to avoid method improvement or innovation, but to avoid the trap, seek real improvement or innovation, and seek more effective improvement or innovation.

Funding

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).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

Bai MY, Ren RE (2001) Research on international comparison and its methodology. *China Soft Sci* 1: 107–113.

Chen MG (2020) Practical Evaluation and Suggestions on China's Participation in the International Comparison Program. *J Guizhou Provincial Party School* 2: 28–34.

Chen Y (1993) Purchasing power parity method and its theoretical defects. *World Econ Politics* 12: 6–10.

Eastwood MA (2017) Heisenberg's uncertainty principle. *QJM: Monthly J Assoc Physicians* 110: 335–336.

Gao MX, Li JP, X J, et al. (2006) *System of National Accounts: Principles and China's Practice*, Beijing: China Renmin University Press.

Jin D (2006) Measuring China's Business Cycles. *Econ Res J* 7: 41–48.

Ferrari G, Laureti T, Mostacci F (2005) Time-Space Harmonization of Consumer Price Indexes in Euro-Zone Countries. *Int Adv Econ Res* 4: 359–378.

Lv GM (2006) Economic cycle fluctuation: measurement method and China's experience analysis, Doctoral thesis, Dongbei University of Finance and Economics.

Ma XJ, Liu XY, Wei XX (2016) The Empirical Research of International Comparison Program Based on Purchasing Power Parity Method and Exchange Rate Method: Take China as an Example. *China Econ Stat Q* 1: 61–73.

Peihua C (2007) The Value Scale to Unite Volume and Value in International Economic Comparison: Analyzing the Shortage of ICP Methodology of UN. *Stat Res* 5: 90–92.

Qiu D (1996) Some Thinking of International Economic Comparison. *Stat Res* 6: 16–21.

Qiu D, et al. (2004) *The Lady Tasting Tea—How Statistics Revolutionized Science in the Twentieth Century*, Translated, Beijing: China Statistics Press.

Qiu D (2018) The Assumption of Pure Price Ratio Hidden in the BH PPP and its Economic Implications. *China Econ Stat Q* 2: 38–67.

Qiu D (2020) The debate between purchasing power parity and market exchange rate in international economic comparison. *China Stat* 4: 4–6.

Wang CQ (1993) Research on several issues of the United Nations International Comparison Project (ICP), Doctoral thesis, Dongbei University of Finance and Economics.

Wang CQ (1995) Empirical Analysis on the deviation of per capita income converted by exchange rate method and ICP method. *Res Financ Econ Issues* 5: 39–44.

Wei HQ (2018) The Mechanism and Improvement of the Summary Method of World Bank International Comparison Project: Comment on Measurement Method of Real Scale of World Economy. *China Econ Stat Q* 2: 245–258.

Mitchell WC (1913) *Business Cycles*, Berkeley: University of California Press, 449–450.

World Bank (2004) *World Development Report 2005: A Better Investment Climate for Everyone,* Washington, D.C.: Copublication of The World Bank and Oxford University Press.

Xu Qiang, Chen HC (2017) Comparison Between International Comparison Program and Eurostat-OECD PPP Program. *Int Econ Rev* 2: 131–143+8.

Xu XC, Liu WQ (2021) Problems Faced by China in Participating the International Comparison Program. *Stat Inf Forum* 36: 3–11.

You SB, Gao Y (2020) Construction Process and Prospect of China's National Economic Accounting System. *Stat Decis* 36: 10–14.

Yu FD (1995) Methods, limitations and improvement measures of PPP for international comparison of GDP. *Stat Res* 1: 67–72.

Yu FD (2003) Analysis of the discrepancies of the world economic sizes based on PPP and exchange rate estimations. *Stat Res* 8: 3–7.

Yu FD (2015) The methodology, the results and the limitations of 2011 ICP round by the World Bank. *Stat Res* 1: 11–19.

Ziebart DA (1987) Parity exchange rate deviations and the misvaluation of reported assets under SFAS No. 70. Working Paper, University of Illinois at Urbana-Champaign.