Abstract

Japanese monetary policy uses the consumer price index as a metric for price stability. Despite a major effort to improve the index, the current state of the Japanese CPI seems to have a large number of deficiencies relative to the US CPI. For example, Japanese “chaining” uses a different formula than US chaining and there is little attention paid to lower-level substitution biases. It is hard for an outside observer not to conclude that there is a substantial upward bias and considerable measurement error in the Japanese CPI.
On March 9, 2006, the Bank of Japan issued a statement clarifying its thinking on price stability: “Price stability is, conceptually, a state where the change in the price index without measurement bias is zero percent. Currently, there seems to be no significant bias in the Japanese consumer price index.” This was an extraordinary statement given that the same report listed eight possible biases in the CPI and scant evidence about their magnitudes. With this statement and the comfort of 5 straight months of positive year-on-year CPI inflation, the BOJ ended its policy of quantitative easing. After four more months of CPI growth of around 0.6 percent per year, the BOJ raised interest rates by 0.25 percent.

But the economic gods are jealous ones, and they don’t take kindly to central bankers believing in flawed statistics. Little more than five months after the BOJ noted the that there was no bias in the Japanese CPI, the Ministry of Internal Affairs and Communications (MIAC) updated the CPI weights (which was just one of the eight possible sources of bias that the BOJ had dismissed) and revealed that inflation had been consistently negative (year on year) for every month until the BOJ raised rates. While inflation was standing at a positive 0.1 percent annual rate in March of 2006, it went negative in April (0.6 percent below the old CPI) following the end of quantitative easing. When the BOJ raised rates to 0.25 percent, inflation was only 0.2 percent, in other words the BOJ prevented real interest rates from going negative.

In this paper, I want to leave aside the question of whether the BOJ would have or should have raised interest rates if they had realized the significant bias in the Japanese CPI. Rather, I would like to discuss what I see as some of the sources of this bias and provide my sense of what I see as the magnitudes in Japan. In this note, I will go over what I think is a useful way of
thinking about price stability and then try to discuss how the Japanese CPI is computed differs from this standard. I will also provide some estimates of the magnitudes of the deviations using import data as well as rely on estimates of these deviations based on existing studies in the US and Japan.

Historically, countries often maintained de facto price stability in terms of one good (e.g. gold or silver), however, as countries gave up specie, monetary authorities lost their simple measure of the value of currency. The use of fiat currency required monetary authorities to have some benchmark against which to measure the “value” of the currency. The problem they face is that it is not feasible to measure all prices in the economy, and even if we could measure all prices, it is not obvious how we should aggregate them. Indeed, it is impossible in general to summarize a vast vector of all price changes in an economy with one number: vectors can’t be summarized by scalars.

This suggests two approaches. The first is to take an atheoretic approach to price measurement. One can construct a price index using an arbitrary – e.g. a Laspeyres index – and let the policy makers do with it what they will. In general, the change in the index will not correspond to rise in expenditure necessary to achieve a certain level of utility in two time periods. This raises questions about the index’s usefulness. Policymakers often have to make important decisions about monetary policy and expenditures based on price information, so ad hoc approaches to price measurement are often not very useful to a policymaker trying to maximize national welfare. Baring computational errors, ad hoc indexes are not wrong, they are just not interesting.

This creates a tension within statistical agencies that want to produce measures of prices that are both correct and meaningful. For example, the MIAC states that Japan’s index is simply
an atheoretic cost of goods index on some portions of its website: “it is necessary to pay attention that the CPI intends to measure the price movements themselves, not to measure movements of living expenses with changes of varieties, qualities or quantities of goods and services.” This is an important, true statement. The CPI will in general fluctuate as prices fluctuate although the prices may not be correlated with changes in the cost of living.

The problem is that if the MIAC is producing indexes of prices that are determined only by the formula used, then the BOJ need not pay attention to value of the CPI when defining “price stability” since the index is not measuring prices in a coherent way. Perhaps the BOJ should define “price stability” as the value of the CPI minus two, or the CPI minus 3 in recessions and minus one in booms. These are all arbitrary mechanisms for defining price stability and one may like one or another based on aesthetic grounds but not based on logical ones.

Nevertheless, it is hard to explain to the public why the government pays so much attention to a meaningless number. In order allay public fears, in the same website in which the MIAC explains that the CPI is a not a measure of living expenses, it claims that it is a measure of these expenses “The index shows changes in the total amount of expenditure required to purchase the equivalent goods and services purchased by households in the base year”. But how can the index both be and not be a measure of living expenses?

The answer is that statistical agencies in Japan and elsewhere have taken a largely meaningless number (the Laspeyres index) and tried to make it track a meaningful one: a superlative index. “Biases” in this context are defined to differences the CPI and a superlative index. The attempts to hold on to the Laspeyres formula while make it perform like a Tornqvist have turned the CPI into a mix of Laspeyres indexes, sample rotations, imputations, and
hedonics to be fed to eager economists and policymakers. It is hard not to read through the 141 pages of Japanese adjustments to the Laspeyres index and not think of the increasingly desperate attempts to rescue Ptolomy’s earth-centric view of the universe with equants and epicycles. At what point do we call the project into question on practical, if not aesthetic grounds?

The obvious alternative to making a Laspeyres index behave like a superlative index is to just use a superlative index. Superlative indexes have the desirable property that they can be derived based on certain assumptions about consumer behavior. For example, most index number theory is based on the assumption that there exists a well-behaved representative indirect utility function that will tell us whether, for a given amount of expenditures, one set of prices will yield higher or lower utility than another set of prices. A critic could reasonably argue that there may not be a representative utility function, but this would constitute more a critique of the entire project of designing a policy to raise social welfare than of the price index itself.

Superlative indexes are vastly simpler functional forms than the consumer price index and have the added property that they can be thought of a second order approximations of an arbitrary utility function. The nice thing about these indexes is that if one accepts the assumption that there is a representative utility function, these indexes approximate how much utility is rising or falling in response to price changes. The major drawback of these indexes is that they require past and current weights. Since current weights are not available, these indexes also require revisions if one approximates the current weights, but these revisions are typically much smaller than those necessary when using a Laspeyres index.
Price Indexes in Japan and the US: Some Important Differences

Overview

Much of the theoretical work for biases in the US CPI has been developed by government and academic economists interested in obtaining better measures of inflation. While the US has resisted officially using utility-theory based indexes, the BLS has implemented a large number of modifications to the CPI that make it perform more closely to this benchmark. Indeed the much of the Boskin Commission’s report was explicitly about eliminating deviations (biases) between the CPI and the Tornqvist superlative index. Although the CPI is nominally a Laspeyres index, in practice the current US CPI contains so many modifications of the original formula (actually developed by William Fleetwood [1707] not Etienne Laspeyres)\(^1\), that in practice it is quite different.

The US has also benefited enormously from the government’s interest in data collection. Ariga [2003] reports that in 2002, the US government spent 10 times more than Japan on the collection of statistics. Ariga documents that number actually understates the difference for a number of reasons. First, 68% of Japanese statisticians are involved in the collection of agricultural statistics (as opposed to 1.5% in the US). Second, the Statistics Bureau of the MIAC (Ministry of Internal Affairs) has only 10 people with a masters level education and no one with a PhD. The US government, by contrast employs 2000 statisticians and economists. While we could not find information on the training level of the statisticians at the BLS, every economist and statistician opening listed on the BLS website (checked on 8/25/06) required the applicant to have at least a masters degree. Moreover every economist and statistician in the BLS over Grade

\(^1\) Diewert, Erwin (1999) “
9, which constitutes the vast majority of these positions, must hold a masters or higher degree. Taken together, the thirty-to-1 spending and manpower differences are likely to generate substantial differences in the sophistication Japanese and US statistical agencies can bring to bear on data collection and processing.

It is therefore not surprising that the Japanese CPI is constructed using different procedures than those in the US. The Japanese CPI is constructed in accordance with the International Labor Organization minimum standards, but this is a low threshold. The typical member of the ILO has a per capita income one twentieth of Japan and can’t be expected to implement the sophisticated price measurements developed in the US. As a result, the methodologies used by Japan and US differ substantially, and one should be very cautious doing cross-country comparisons of aggregate prices. To understand these differences, one needs to delve a bit deeper into the statistics.

The US CPI contains two levels of aggregation. At the upper level, 211 strata level price indexes in each of 38 areas (or regions) are combined either using a Laspeyres formula in the case of the standard CPI or a Tornqvist formula when using the chained CPI. This means that there are 8018 item-area indexes that are aggregated at the upper level. Each of these indexes is in turn based on a lower level sample of approximately ten price quotations per item-area (85,000 price quotes all). This lower level price quotation is critical for the BLS’s approach to price measurement. Almost all of the deviations from the standard Laspeyres index in the US case – hedonics, geometric averaging of prices, sample rotations, etc. – occurs at this lower level. Thus while the upper level of the US CPI is Laspeyres, the lower level is a complex combination of geometric averages, hedonics, and imputations. Since much of the substitution by consumers

---

2 Some of these strata are comprised of 2 or more Entry level items to account for different items within a strata. The 305 ELIs are aggregated into to the 211 strata in the CPI.
is done at the lower level (among, say, different brands of the same good), than across different expenditure classes (e.g. cars vs. televisions), this two tiered approach corrects much of the substitution bias.

The Japanese CPI, by contrast, is much closer to a pure Laspeyres index. At the upper level, the Japanese CPI is more disaggregated than the US CPI. Japan uses 598 items in its CPI instead of the 211 strata used in the US, and surveys these prices in 167 municipalities across Japan as opposed to the 38 in the US. Thus, at the upper level Japan has more than ten times the number of price series in its CPI calculation. However, there is a big cost to the significantly larger and more geographically dispersed sample: the lower level of the Japanese CPI is virtually non-existent. Moreover most of the research on the Japanese CPI, with the notable exception of Ariga [2003]) has focused on issues at the upper level biases, leaving most of what the US has focused on untouched. For example, the Japanese don’t use geometric averages and only do hedonic regressions on personal computers and most recently digital cameras as opposed to the vast array goods analyzed this way in the US.

Lower Level Substitution Bias

The Japanese CPI, for example, typically only uses one price quotation per item-area as opposed to 10 in the US (Ariga [2003]). One of the major drawbacks of this procedure is that the Japanese CPI cannot make any adjustments for what the BLS calls “lower-level substitution”: the ability of consumers to switch away from high priced bands to low priced brands of a particular item in a particular region. With only one price per item area, the Japanese cannot implement one of the most important adjustments to the CPI in the US: geometric averaging at the lower level. This has been implemented in the US since 1999. The advantage of geometric
Averageing is that it allows for substitutability among the various goods that make up an item index. The US formulation is equivalent to assuming that goods have an elasticity of substitution of unit. Shapiro and Wilcox [1997] document that the geometric mean of prices is a close approximation of the superlative Tornqvist aggregator. This is not that surprising since the geometric average is a first order log approximation of an arbitrary utility function while the Tornqvist can be thought of as a second order approximation.

Sample Rotation

MIAC (MIAC [2006] and Bureau of Labor Statistics [2006]) uses a far simpler sampling procedure than the BLS in the construction of the lower level sample. The Japanese approach to sampling of goods follows a top-down approach. Periodically, MIAC decides what items to sample and then adds them to the CPI. So, for example, one simply could not have done hedonic regressions on computers in Japan prior to 2000 because computers were not an element of the Japanese CPI. By contrast, the Japanese CPI was still tracking abacus lesson fees as late as 1999. Similarly CD players and mp3 players are not included but phonographs are (Shiratsuka [2006]). To the extent that the prices of computers, CD players, and mp3 players have fallen faster than the prices of abacus lessons and phonographs, one might suspect an upward bias in the CPI due to the slow rotation of items.

By contrast, the BLS uses much broader definitions of its strata and then samples products based on what consumers say they are purchasing. For example, the BLS will interview consumers about what audio equipment they buy (as home and car audio is a strata), and then use the responses to decide whether consumers in 2005 are spending more money on CD players,
mp3 players, phonographs, or whatever the person responds. Once they’ve ascertained the product that consumers are now purchasing, they then proceed to the store to price that particular item.

These different procedures produce very different rates of product rotation. The BLS schedules a rotation of 25 percent of its sample every year based on these interviews and achieves full sample rotation every 4 years. While clear documentation of the rate of lower level rotation is hard to find, Japan’s items (e.g. personal computers and phonograph players) appear to have a scheduled rotation rate of about 10 percent every 5 years. Thus scheduled rotations in Japan occur at one tenth the rate of that of the US. Indeed the rate of product rotation is more comparable to the way the US CPI was constructed prior to 1981. Concerns over failure of the CPI to rotate its sample fast enough (particularly the location of purchases) resulted in a major revision of BLS procedures to correct these problems in 1998.

A separate but related problem in the Japanese CPI is that it uses a non-random sample of products. In the US CPI, interviewers ask households for representative products in each strata and then use sophisticated sales-weighted random sampling within each store to identify which product will be selected off a shelf. The Japanese CPI is based on surveys of prices of the largest seller. In simulations on US consumption data and Japanese import data (the Japanese import price index) done by Broda and Weinstein (in process), we find that non-random sampling leads to extremely large measurement errors in a price index. These measurement errors arise because there is much less price volatility for the largest goods than for all goods.

---

3 The ten percent number was based on the 71 new items included in the 2000 revision of the CPI and 55 items eliminated out of a total of 598 items. Neither set of numbers includes forced rotations arising from the appearance and disappearance of goods. Shiratsuka [2006] reports that the MIAC reviews items and potentially changes them every 2.5 years.
Chaining

Both Japan and the US now report a “chained” CPI, but the methodology for chaining is completely different in the two countries. The US chained CPI (the C-CPI) uses a superlative index number formula at the upper level and geometric price averaging at the lower level. This means that the US chained CPI can be thought of as somewhere between a first and second order approximation of an arbitrary utility function.

Japanese “chaining” is actually what the US calls “annual weight updating”. The BLS now updates the base weights in the CPI every two years. While more frequent weight updating may reduce substitution biases, it is not sufficient to eliminate them. To understand why, think about how the problem arises. One of problems in Laspeyres indexes is that persistence in price movements can cause the base weights to deviate significantly from the current expenditure shares. Earlier versions of the CPI suffered from this problem as often the weights were updated often less than once a decade. If prices in some sectors are trending upward, then the Laspeyres index will overweight those sectors because it will not adjust for the fact that consumers will buy a relatively fewer items if the prices of those goods are rising relative to other goods.

Increasing the frequency of weight updating is conceptually quite different from adopting a superlative index. Regardless of the frequency of the weight updating, one cannot change the fact that the Laspeyres index uses historical weights and consumption decisions today are going to be driven by current weights. As long as relative prices have changed the index will differ. Thus while the substitution bias may be mitigated by the fact that the MIAC now updates the CPI’s base weights more frequently, they have not corrected for the fact that if prices change, the base weights will not be the same as the current weights. For goods whose prices are changing
rapidly (e.g. high tech goods), the Japanese “chained” CPI will not correct for the substitution bias.

Estimates of Bias

In a very careful and controversial paper, Shiratsuka [1999] estimated that there was an upward bias of 0.9 percent per year in the Japanese CPI. In later work, Shiratsuka [2005, 2006] claims that the revisions in the CPI have largely eliminated the bias. One must be careful about this interpretation. First, it is important to recognize that the papers refer to two different biases. Shiratsuka [1999] estimated the impact of upper level substitution bias arising from using a Laspeyres index and found that to be small. This is not that surprising as the ability of consumers to substitute between categories like housing and transportation may be much smaller than among types of housing or cars. Indeed, the Boskin Commission’s estimate of the upper level substitution bias for the US of 0.15 percent per year is very close the average Shiratsuka’s high and low estimate for Japan of 0.13 percent.

Shiratsuka [2005 and 2006] estimates the differences between a fixed weight Laspeyres, a Japanese-style chained Laspeyres index, and the midpoint weighted Laspeyres (which uses weights between the base year and the current year) and finds that they all produce similar results. But here is the problem. None of the indexes that Shiratsuka examines are superlative indexes. This means that none of them completely correct for the substitution bias. But if we know that the fixed weight Laspeyres has a significant substitution bias and that all of the MIAC’s “corrections” to this index produce similar results, then the only logical conclusion is that these “corrections” failed to eliminate the bias. This is probably not that surprising since none of the indexes used by the Japanese government chain in the US sense.
Nevertheless, these studies are still focused on biases at the upper level, which even the Boskin commission found to be small. A much bigger question concerns the lower level substitution bias. In the US, the quantitative importance of this adjustment has been estimated to quite large. Moulton and Smedley [1995] estimated that the switch to geometric weighting at the lower level biases the CPI upward by 0.5 percent per year. Shiratsuka argues that the lower level substitution bias is likely to be low because inflation is low. This argument seems problematic because the substitution bias is not driven by the aggregate inflation rate, but rather by deviations in relative prices. To put this simply, if all prices in the Japanese economy rose by 3 percent, there would not be any substitution bias since consumers would have no incentive to switch expenditures across categories. The substitution bias arises from variation in relative prices, it is not obvious that these variations are higher as inflation ranges from 0-3 percent.

Third, Shiratsuka argues that the fact that Japanese item-strata are more disaggregated than those in the US compensates for the lower-level bias. This seems unlikely. The key point to bear in mind is that the lower-level substitution bias is a formula bias arising from the fact that the Laspeyres index does not allow for substitution at the lower level. This formula bias will be present regardless of the level of aggregation. Put simply, a Laspeyres index would exhibit formula bias even if one observed every price in the Japanese economy. If you use the wrong formula, you will get the wrong answer (unless you are very lucky).

To the extent that we can apply US estimates of these biases to Japanese data, it appears that there are substantial biases still to be corrected. Upper and lower level substitution biases are likely to amount to 0.6 percent per year with the vast majority of this number coming from the lower level bias. Moreover Reinsdorf [1993] estimated that that biases in US CPI arising

---

4 Greenlees argues the lower level bias in the US CPI prior to the introduction of geometric averages are 0.2 percentage points per year but, as the Boskin Commission documents, this lower number reflects the remaining bias after the BLS implemented procedures to reduce the lower-level substitution bias.
from not adjusting for new points of sale amounts to 0.25 to 0.4 percent per year. New goods and quality upgrading is likely to add another 0.6 percent per year to this. This suggests a total bias of 1.3 percent to 1.6 percent.

Estimates of the bias based on Japanese data can also total fairly large amounts. Ariga [2003] examined the difference between point of sale price data and the prices reported in the CPI. He consistently found that the CPI overstated inflation by 1.5 percent to 2 percent per year. This suggests that it is not hard to find substantial biases in the reported indexes. He later generated his “conservative” estimate by taking these numbers and dividing them by 3, but it is not clear to me. Certainly the available data seems to indicating a bias that is in line with the bias in the US.

Finally, in work in process, Christian Broda and I are computing the bias in the import price index. This is a good index to focus on because the availability of 9-digit data means that we can replicate the BOJ methodology for computing its price index (which is similar to the Japanese CPI methodology), and then recompute the index using the full set of prices. Our preliminary results indicate that there is a 1.5 percent upward formula bias in using a the Laspeyres index relative to a Tornqvist.

Conclusion

Japanese policymaking is hampered by inadequate statistical resources. Even in 2006, measurement errors have resulted in the central bank thinking that there was inflation when in fact there was deflation. This can have potentially serious consequences for monetary policy. Moreover, with social security and other government payments indexed to the CPI, an upward bias in the CPI can have serious fiscal impacts.
In this note, I have discussed a few of the many ways in which an upward bias can creep into the CPI. These biases seem to lead to an upward bias in the official number of around 1.5 percent. Importantly, I have not even touched additional biases such as outlet substitution, quality upgrading, and new goods, which are likely to push this number up substantially. A clear policy implication is that more resources need to be devoted to the collection of Japanese price statistics and that central banks that want to avoid deflation should take these biases into account.
References (Incomplete)
