How Much Do Official Price Indexes Tell Us About Inflation?*

Jessie Handbury
University of Pennsylvania

Tsutomu Watanabe
University of Tokyo

David E. Weinstein
Columbia University and NBER

December 10, 2015

Abstract

Official price indexes, such as the CPI, are imperfect indicators of inflation because they employ *ad hoc* formulae. These differ from the theoretically well-founded inflation indexes favored by economists. This paper provides the first estimate of how accurately the CPI informs us about "true" inflation, as measured by a superlative inflation index. Our comparison of this true inflation index with the CPI indicates that the CPI bias is not constant but depends on the level of inflation. When measured inflation is low (less than 2.4% per year) the CPI is a poor predictor of true inflation even over 12-month periods.

*We would like to thank Morgan Hardy, Scott Marchi, and Molly Schnell for excellent research support. We also would like to thank Sakai Ando, Kosuke Aoki, Erwin Diewert, Kevin Fox, John Greenlees, Masahiro Higo, Emi Nakamura, Hugh Patrick, and Marshall Reinsdorf for excellent comments. David Weinstein would like to thank the National Science Foundation (SES-11227493) and the Center on Japanese Economy and Business for generous financial support. Jessie Handbury would like to thank the Research Sponsors' Program of the Zell-Lurie Real Estate Center and the Wharton Global Initiatives for financial support. Tsutomu Watanabe would like to thank the Japan Society for the Promotion of Science for financial support (JSPS Grant-in-Aid for Scientific Research no. 24223003). Some results are calculated based on data from The Nielsen Company (U.S.), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

1 Introduction

Although we have long known that the price indexes constructed by statistical agencies, such as the Consumer Price Index (CPI) and the Personal Consumption Expenditure (PCE) deflator, measure inflation with error, most economists treat a change in an official price index as equal to a change in inflation. A theoretical macroeconomics literature starting with Svensson and Woodford [2003] and Aoki [2003] has noted that stochastic measurement errors imply that one cannot assume that true inflation equals the CPI less some bias term. In general, the relationship is more complex, but what is it?

This paper provides the first answer to this question by analyzing one of the largest datasets ever utilized in economics: 5 billion Japanese price and quantity observations collected over a 23 year period. The results are disturbing. We show that when the Japanese CPI measures inflation as low (below 2.4 percent in our baseline estimates) there is little relation between measured inflation and actual inflation. Outside of this range, measured inflation changes are correlated with, but understate, actual inflation changes. In other words, one can infer inflation changes from CPI changes when the CPI is high, but not when the CPI close to zero. This non-linear relationship between measured and actual inflation has important implications for the conduct of monetary policy in low inflation regimes, especially in countries that use Dutot indexes, such as Belgium, Germany, Japan, and the U.K. We also show that under the U.S. procedure of geometric averaging of prices, the non-linearity and biases would be reduced but not eliminated.

The basic intuition is straightforward. There is little disagreement that superlative price indexes, such as the Törnqvist, have the best theoretical properties of any price index valid for understanding homothetic preferences [Diewert and Nakamura, 1993]. However, the construction of such indexes requires detailed price and quantity data, which, until the invention of scanner technology, was too expensive to collect on a systematic basis. Thus, official price indexes in all countries are constructed using other functional forms. We demonstrate that the measurement error that results from using the wrong formula is not constant over time;

it is instead noisy. True inflation is also volatile, and the empirical macroeconomics literature has documented that the volatility of inflation rises with the level of inflation (see, e.g., Okun [1971], Friedman [1977], Taylor [1981], Ball et al. [1988], Ball and Cecchetti [1990], Kiley [2000]).[1] We study the volatility of *both* true inflation and the CPI measurement errors with surprising results. We confirm that the volatility of true inflation rises rapidly with its level, but we also find that the variance of the CPI measurement errors does not. Thus, when inflation is high, most of the observed movement in the CPI is due to actual inflation movements, but when the inflation is low, much of the movement in the CPI is noise.

To understand the math underlying this result, think of the CPI as having two components: a signal of true inflation and a measurement error, or noise. If the measurement errors are uncorrelated with inflation rates, it is easy to show that the CPI inflation will be more informative about actual inflation if the signal-to-noise ratio is high. If this ratio is high, the variance of true inflation must be a large component of CPI variation, and one should expect that true inflation moves nearly one to one with the CPI. If the variance of true inflation is small relative the movements in the CPI, then the signal-to-noise ratio will be low, and it is safe to largely ignore CPI movements when trying to determine inflation changes.

The indexes used to construct the CPI, such as the Laspeyres, were originally developed to measure fairly substantial price changes—differences in the cost of living over a 250 year period— rather than the month-to-month or year-to-year variation the CPI measures today (Diewert [1993]). Indeed, our results suggest that the choice of index number doesn't matter much for measuring large values of inflation. However, as global inflation rates have fallen

[1] Friedman [1977] argues that "a burst of inflation produces strong pressure to counter it. Policy goes from one direction to the other, encouraging wide variation in the actual and anticipated rate of inflation." A similar idea is proposed by Ball [1992], who argues that when inflation is low, there is a consensus that it should be kept low; however, when inflation is high, there is disagreement about the importance of reducing it, leading to high variability of inflation. Taylor [1981] suggests that accommodative monetary policies may lead to high inflation and greater variability in response to supply shocks. Cukierman and Meltzer [1986] argue that an exogenous increase in the variance of monetary control errors gives a central bank a stronger incentive to create surprise inflation, leading to a high inflation and high volatility of inflation. Ball et al. [1988] show, using a menu cost model, that high trend inflation reduces nominal price rigidity and thus steepens the short-run Phillips curve. As a result, shocks to aggregate demand have smaller effects on output but larger effects on inflation. Gagnon [2009] confirms that the micro price data is consistent with this story by showing that the frequency of price adjustment increases with the level of inflation.

close to zero in many countries and central banks and market participants interpret inflation movements of one percent or less as significant, the precise formulas and approximations used to measure inflation start to matter much more. Just as a bathroom scale is not appropriate for measuring the weight of a letter, index numbers that may be very good at distinguishing between inflation rates of 5 and 10 percent may not be good at distinguishing inflation rates of 1 to 2 percent. Indeed, our results indicate that there is a much tighter relationship between the CPI and inflation in high inflation regimes than in low ones.

This has profound implications for how one should think about inflation and monetary policy. First, our results suggest that in general *one cannot write true inflation as a linear function of the CPI*. In particular, true inflation cannot be assumed to equal the CPI less a constant bias term. These non-linearities matter enormously in our data. For example, in our baseline estimates based on Japanese data that we will discuss later, we find that when the annual Japanese CPI registers an inflation rate of 0 percent, true inflation is -0.8 percent—but when CPI inflation is 2 percent the bias rises to -1.8 percent. In other words, a 2 percent CPI inflation target is actually a price stability target, when using annual data. Since the methodology underlying the Japanese CPI is more prevalent internationally than that used in the U.S. PCE deflator, this potentially has broad implications.

A second important implication concerns how central banks and other economic agents should respond to CPI inflation movements. Again, the non-linearity arising from the fact that the signal-to-noise ratio in the Japanese CPI is much higher in high-inflation regimes matters. We estimate that a one percentage point rise in the CPI should raise one's assessment of true inflation by only half a percentage point when inflation is less than 2 percent, but this same increase should raise one's assessment by 2 percent when inflation is high. In other words, failure to take into account the change in an official price index's signal-to-noise ratio is likely to mean that central banks pay far too much attention to official price indexes when they are low and not enough attention when they are high.

Our results also have implications for the U.S. PCE deflator. Although we can show that the PCE methodology results in substantially smaller biases than the Japanese methodology, similar patterns remain. When we replicate the PCE deflator methodology using Japanese data, we find the PCE deflator methodology overstates inflation by about 0.6-0.7 percentage points per year when inflation is in the 1-2 percent range and overstates inflation changes by 20 percent when inflation is below 1.75 percent.[2]

Finally, our results call into question whether our economic statistics are sufficiently well-calibrated to do cross-country comparisons of economic performance among advanced countries. For example, the BLS estimated that when the U.S. switched to geometric averaging of prices in 1999, it reduced measured inflation by 0.34 percentage points per year.[3] Mechanically, this change also must have raised U.S. real GDP and wage growth by the same amount since real growth rates are just nominal ones less the change in the price index. We should expect that countries like Japan that did not make this change, therefore, would have reported lower *measured* inflation and higher growth rates if they had made this change. Indeed, we find that properly measured prices fell much faster in Japan than what the Japanese CPI recorded. If the results for our subsample extend to the full sample of prices, this meant that Japan grew 0.8 percent per year faster between 1993 and 2010 than official statistics suggest. Maybe part of the reason why Japan's growth rates have appeared so low is due to how they were measured.

In order to demonstrate these points, our paper makes use of the Nikkei Point of Sale database.[4] Nikkei collects daily scanner purchase data from hundreds of large and small stores covering hundreds of thousands of bar codes spread across Japan in the grocery sector. These data are typically sold by Nikkei to customers interested in accurately measuring the sales of individual bar codes. Our data covers the period from 1988 to 2010 and contains 4.82 billion price *and* quantity observations. The Japanese CPI, like the CPIs of most countries, contains no quantity data and a tiny fraction of the number of price observations. Crucially,

[2]Most prior work in this area has focused on the level, as opposed to the variability, of the bias (see the excellent surveys by Hausman [2003], Lebow and Rudd [2003], and Reinsdorf and Triplett [2009]).

[3]http://www.bls.gov/cpi/cpigmrp.htm

[4]This is an expanded version of the dataset used in Abe and Tonogi [2010].

the Nikkei data come as close as one could come empirically to observing the universe of grocery price *and* quantity observations in the Japanese economy. Moreover, the long time span means that we observe Japanese periods of substantial inflation and deflation. These features of the data enables us to construct the theoretically well-founded Törnqvist index of "true" inflation for a set of items covering 17 percent of Japanese consumer expenditure over a period spanning more than two decades.

These data let us make a number of important empirical findings. The first is the volatility in the mismeasurement of the CPI. Our estimates of this error based on grocery data are likely to be substantially lower than for the rest of the CPI because it is much easier to measure the prices of the goods sold in a grocery store than major CPI components like housing, transport and communication, and recreation. Nonetheless, we find that inflation mismeasurement is large. Our results indicate that over 12-month periods, the grocery CPI sometimes overstated inflation relative to a Törnqvist by as much as four percentage points while at other times it understated inflation by as much as 2 percentage points. The fact that the standard deviation in the bias is 0.9 percent means that much of the fluctuation in observed inflation during this time period was due to CPI errors. It is the surprising magnitude of these errors that drive our main finding that the CPI is not very informative when the variance of inflation is low. Second, we show that while there is some tendency for the variance of CPI noise to rise as inflation rises, the rise is small compared to the tremendous increase in inflation volatility that occurs as inflation rises. In other words, while Cecchetti [1997] and Shapiro and Wilcox [1996] were right to point the tendency for CPI noise to rise with inflation because lower-level substitution bias rises with inflation, this effect is very small compared to the rise in inflation volatility.

Because we observe the underlying individual price and quantity data, we are able to be precise about how the non-linear relationship between the two series is generated. We decompose inflation volatility into two components: one related to movements in components of inflation that are common across products and another related to the relative price dispersion across products. Consistent with previous work, we find that relative price dispersion rises with inflation (c.f. Vining and Elwertowski [1976], Parks [1978], Fischer [1981] and Stockton [1988]). The variance in relative price shocks more than doubles as inflation rises past 2.4 percent, but this dispersion explains only a small component of the increase in aggregate price volatility. Most of this increase is explained by the variance in the aggregate or common component of inflation, which rises by a massive 320 percent. The non-linear relationship between inflation and the CPI is primarily driven by the fact that CPI measurement errors are much less sensitive to inflation than either component of aggregate price volatility: the variance of upper- and lower-level measurement errors only increase by 35–60 percent as we move from low inflation to high inflation regimes. The critical factors driving the non-linearity are the higher inflation variance in high inflation regimes and the relative insensitivity of CPI measurement errors to inflation movements.

Our micro level analysis has some practical implications for the construction of consumer price indexes. First our results suggest that upper-level weighting errors—which in the U.S. account for most of the difference between chained and unchained indexes—matter for the average bias but are relatively unimportant in understanding movements in measurement error. Second, we find that one major source of formula error comes from using arithmetic, rather than geometric, averages of prices. This suggests that the move to geometric averaging of prices not only reduces the upward bias in inflation indexes but also reduces the noise in these indexes. Third, we are able to show that geometric averaging is not a complete solution in itself. In our simulations using Japanese data, we show that the PCE deflator methodology still produces significant biases and measurement error relative to the Törnqvist due to the choice of simple geometric averaging rather than weighted averaging. Interestingly, sampling errors seem to be relatively unimportant in understanding why our price indexes are flawed.

The structure of the paper is as follows. Section 2 describes our data and provides a preview of the CPI measurement issues, biases and data challenges. Section 3 describes the econometric strategies we follow in decomposing the relationship between true and measured inflation. Section 4 presents our main empirical findings. Section 5 considers how robust our results are to alternative methods of computing the CPI and alternative datasets. Section 6 examines a number of possible fixes for the CPI measurement problem, and Section 7 concludes.

2 Data

Our paper makes use of two principal datasets. The first is detailed Japanese CPI data. We obtained the price indexes and CPI weights from various issues of the Annual Report on the Consumer Price Index produced by the Statistics Bureau (SBJ).[5] We then used these data to build the CPI for grocery items following the same formulas used by the SBJ. In all of our work, the inflation rate we work with is the inflation rate in a given month relative to the same month in the previous year. This lets us avoid seasonality problems.

The SBJ produces a CPI index conforming to the International Labor Organization standard. The ILO methodology is the basic approach used in all advanced countries.[6] We will discuss differences between the Japanese methodology and that of the U.S. in Section 5.2, but it is worth noting here that the ILO reports that more countries use a methodology akin to the SBJ than that used by the Bureau of Labor Statistics (BLS).[7] Thus, many of the strengths and weaknesses of the Japanese CPI are common to most advanced countries.

Nevertheless, it is worth describing how the index is constructed. Price quotes for each good are collected each month during the span of the same week in stores in hundreds of locations across Japan. Indeed the geographic representation of the Japanese CPI is superior to that of the U.S. CPI. Like the U.S. CPI, there is highly imperfect information available on what weights to use at the lower level. The SBJ defines product type specifications for each

[5]Although the official name of the bureau is the "Statistics Bureau," we will follow the Japanese website and abbreviate it with SBJ, which stands for "Statistics Bureau of Japan." Data are available at the SBJ website (http://www.stat.go.jp/english/data/cpi/). Issues prior to 2001 are only available in paper form.

[6]See http://www.stat.go.jp/english/data/cpi/1585.htm#B2

[7]See `http://www.ilo.org/public/english/bureau/stat/download/cpi/survey.pdf`.

of the item categories and only collects prices for products fitting these specifications.[8] For example, the price quotes for the "Butter" item include only the prices of products that are 200 grams and packed in a paper container, excluding unsalted varieties [Imai et al., 2012]. The SBJ then calculates inflation for each item as a Dutot index; that is, the ratio of the (unweighted) average price quotes collected for an item in one period relative to another. Once these item indexes are constructed, the SBJ aggregates the data to form an overall CPI using "upper-level" weights from its consumer expenditure survey. The upper-level item weights are based on item expenditure shares in a base year that is updated every five years.[9]

The items in the grocery elements of the CPI account for 17 percent of the full CPI. Grocery items are extremely standardized items that are easy to price, so it is far easier to measure price changes for this sample of the CPI than it is for most of the remainder. For example, it is easier to measure the the 30-day price change of a 300 ml can of Coca Cola sold in a particular store than the 30-day price change of other major expenditure items like imputed rent or recreational services. One should therefore anticipate that our paper understates the magnitude of CPI measurement error.

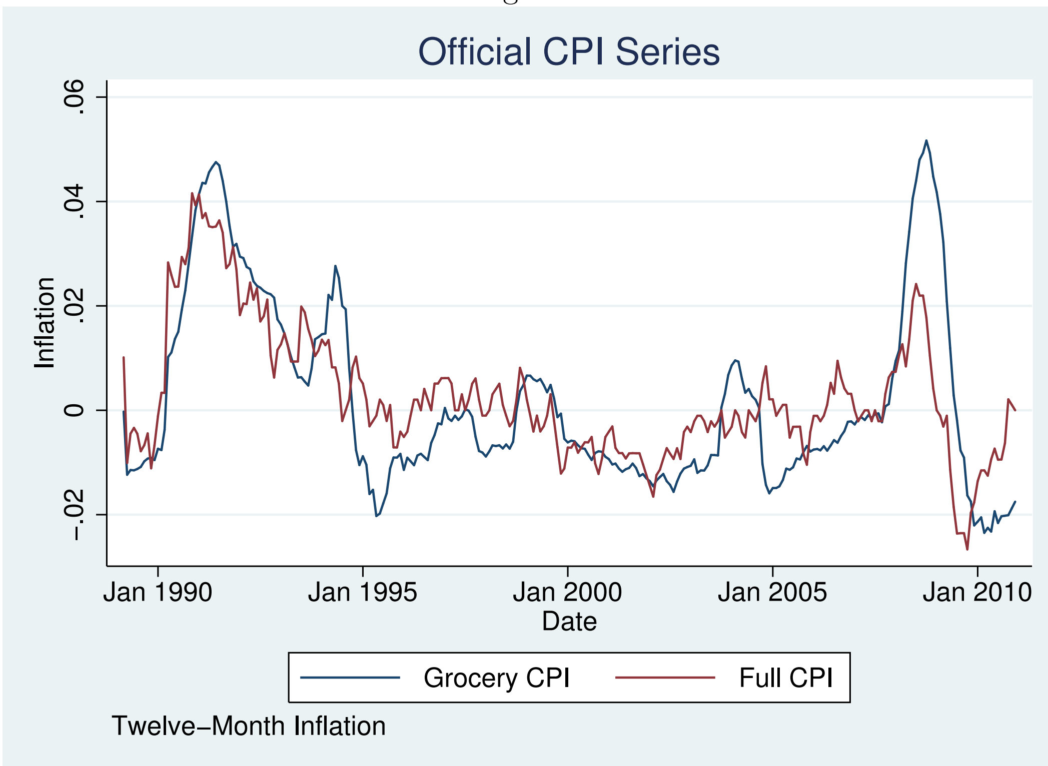
An obvious concern with our focus on grocery items is that this sample of goods might not be representative. While it is true that inflation measurement errors are likely to be much smaller in our sample of grocery items than for the whole CPI, the grocery CPI tends to track the overall Japanese CPI quite closely. Figure 1 plots both the Grocery CPI and the full CPI, and it is immediately apparent that the two are fairly closely correlated ($\rho = 0.8$) despite the fact that grocery items make up less 20 percent of the CPI. Our sample of products captures most of the variation in the official CPI including the overall trend in inflation.

Our second dataset is the Nikkei Point of Sale (Nikkei POS) dataset. The Nikkei data

[8] See Imai et al. [2012] for a more detailed description of this "purposive" sampling rule and a comparison of its performance relative to the probability sampling used by the BLS.

[9] There are a range of differences between how the BLS calculates the U.S. PCE and the Japanese methodology described above. We explore how these technical differences potentially reduce the variance of the errors that drive our main results in Section 5.2 below.

Figure 1



differs from other bar-code datasets in frequency, scope, and length. Nikkei collects data at a daily frequency. We observe the number of units of a bar-coded good purchased in a store on a day and the store's sales revenue for that bar code on that day. In a typical month we observe prices for about a quarter of a million different grocery products, each identified by a bar code. Nikkei also has impressive scope relative to other datasets. Where other datasets focus on one store or chain, the Nikkei POS database includes observations taken from hundreds of grocery and convenience stores spread across Japan each day. This is done so that Nikkei can provide customers with an accurate picture of how goods are sold in general and across different markets. Finally, the truly amazing feature of the Nikkei POS data is its time dimension. Our data makes use of 23 years of these data collected from 1988 through 2010, all together comprising close to 5 billion observations.

The detail in the Nikkei data allows us to construct a price index under fewer assumptions than the official CPI. We measure "true" inflation using a Törnqvist index. As a superlative index, the Törnqvist is a second order approximation of *any* twice-differentiable homothetic expenditure function, and is as close as we can come to computing an exact inflation index without actually specifying preferences.[10] Since our data permits us to compare the prices of identical goods purchased in the the same store using the correct weights for that store, we are able to exactly produce this index.[11]

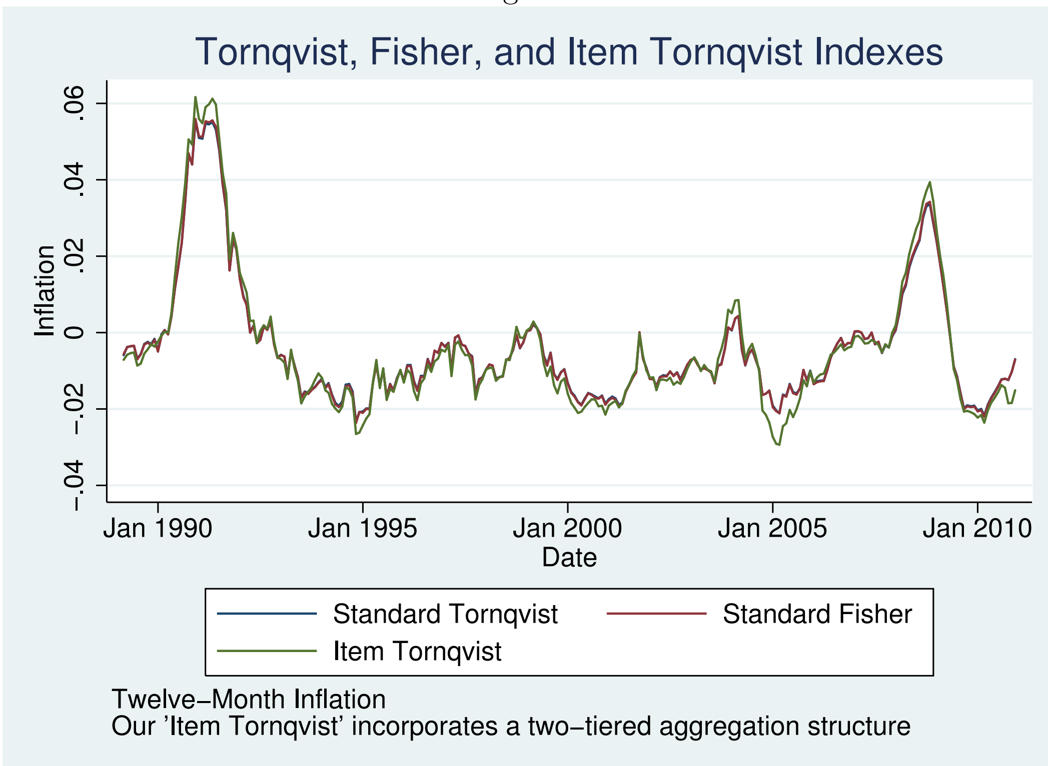
It is difficult to formally compare the properties of the two-tiered CPI with those of the Törnqvist, which does not have a two-tier system. For example, it is not possible to formally assess the importance of measurement error at the CPI's lower level because the Törnqvist has no lower level. In order to make progress on understanding what causes CPI errors, we utilize a Törnqvist index that has a two-tiered structure. We call this two-tiered index the "item Törnqvist" because the lower tier indexes are calculated at the "item" level to correspond with the Japanese CPI and then aggregated using Törnqvist weights. Fortunately, the item Törnqvist produces inflation rates that are almost identical to those of the standard Törnqvist as is shown in Figure 2. Thus, we can be confident that virtually all of the difference between Törnqvist and the CPI can be understood in terms of the decomposition between an item Törnqvist and the CPI and not as differences between the item Törnqvist and the standard Törnqvist.[12] In the interest of simplicity, we will therefore ignore the difference between the item-Törnqvist and the Törnqvist indexes of inflation and refer simply to to the "item Törnqvist" as the "Törnqvist" index throughout. Figure 2 also demonstrates that the single-tiered Törnqvist index is indistinguishable from a Fisher index, an arithmetic superlative index that is also used to measure true inflation.

[10] The Törnqvist measures inflation in common goods prices. Assumptions on preferences are required in order to account for the welfare effects of entry and exit of goods (see, for example, Broda and Weinstein [2010]).

[11] See Appendix B for details on the Törnqvist index calculation.

[12] Diewert [1978] shows that a superlative index, such as the Törnqvist, is approximately consistent in aggregation (that is, the value of an index calculated in two stages approximately coincides with the value of an index calculated in a single stage). It is not surprising, therefore, that all of our results are qualitatively identical whether we use the standard or the item Törnqvist.

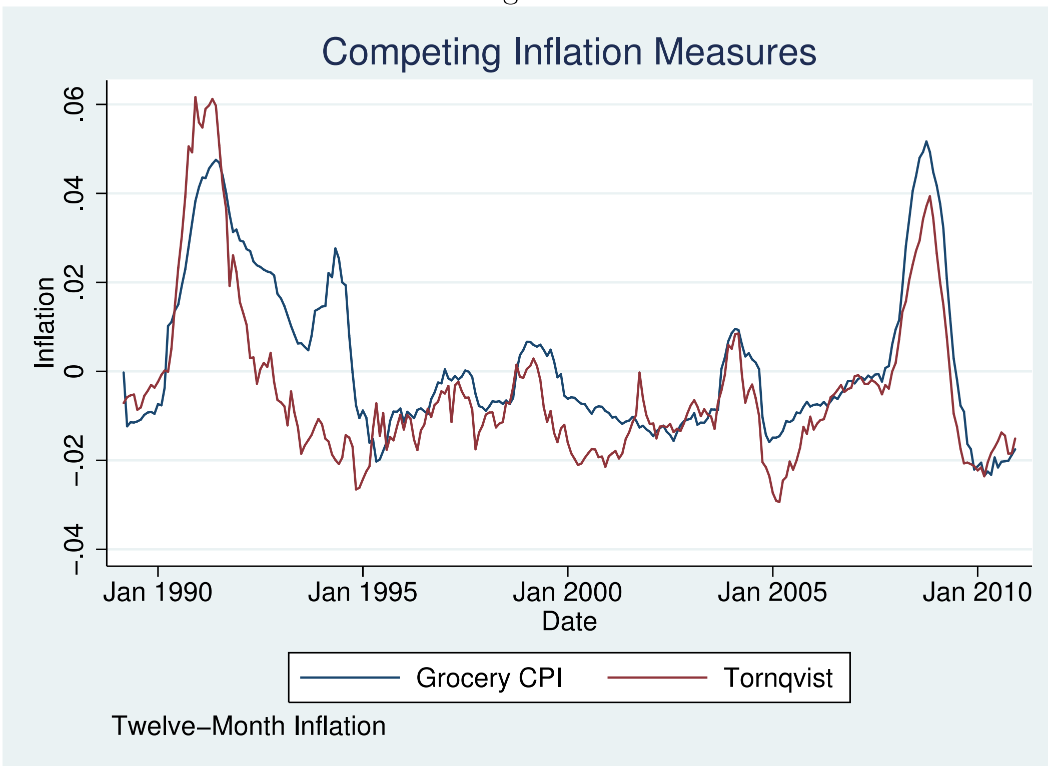
Figure 2



2.1 Measurement Error in the Official CPI

We calculate the measurement error in the official CPI by comparing it with the Törnqvist index. Figure 3 presents the grocery component of the official CPI and the Törnqvist index calculated using the Nikkei data. As one can see in the figure, there are substantial differences between the CPI and the Törnqvist index. Figure 3 suggests that there is a very clear upward bias in the official index. While the Törnqvist index became negative in 1993, the official index did not register deflation until 1995. More generally, the correlation between the CPI and the Törnqvist is much less tight than what one might expect for two indexes measuring the same component of inflation. The correlation between the two indexes is 0.87, which reflects the fact that there are some periods in with the two indexes differ quite substantially. For example, in 1994 the CPI was registering an inflation rate that was at one point 4.7 percentage points above that of Törnqvist, in 1991 CPI *understated* inflation by almost 2 percentage points, and between 1995-1999, the bias was close to zero. Thus, Figure

Figure 3



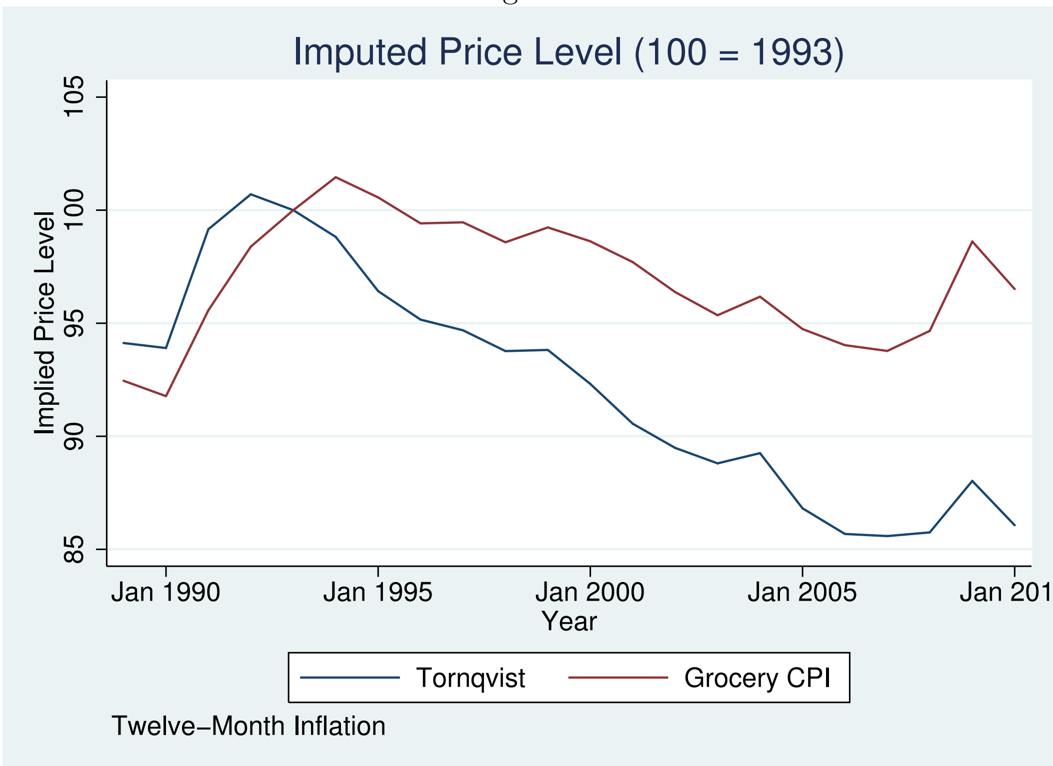
3 indicates that one should be very wary of assuming that one can infer inflation by looking at an official index and subtracting off a constant bias term.

In Table 1, we study the bias in the CPI, defined as the difference between the CPI and the Törnqvist. As we can see in the first column, this difference averages to 0.63 percent per year. The Törnqvist index suggests that Japanese grocery goods entered a period of sustained deflation in 1993. If we focus our attention on that period we see that the official bias in the CPI was actually larger, at 0.76 percent per year.

Table 1
Index Biases

	$\pi^{\text{CPI}} - \pi^{T}$
Annualized Total Bias	0.625
Standard Deviation of Bias	0.961
Annualized Total Bias (Post-93)	0.762
Standard Deviation of Bias (Post-93)	0.763

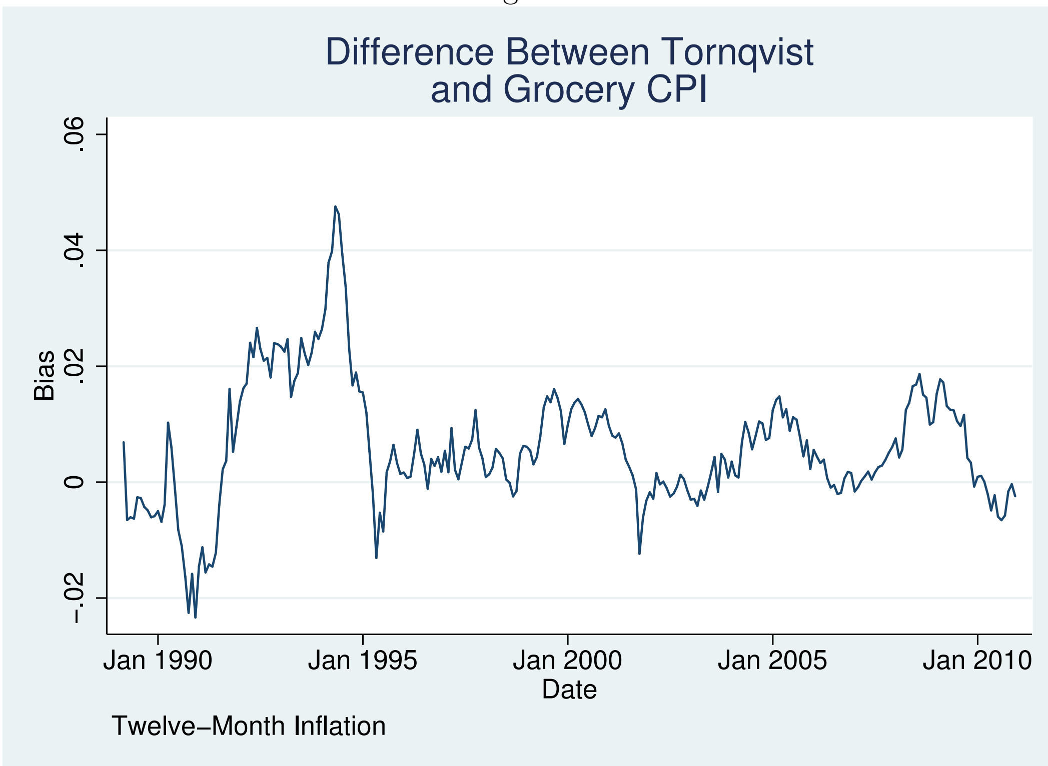
Figure 4



This bias implies a radically different impression of what has happened to the price level in Japan (at least for grocery goods). As one can see in Figure 4 the official price index suggests that prices only fell by 4.2 percent between 1993 and 2010, suggesting that deflation was quite mild. The superlative index, on the other hand, indicates that deflation averaged .91 percent per year, implying a far more substantial 14.5 percent drop in the price level over the whole period. An interesting corollary of this result is that if Japanese CPI inflation overstated true inflation by about one percent a year, then Japanese real consumption growth was also higher by this amount. This suggests that Japan's "lost decade" might not have been nearly as severe as official statistics suggest.

What is most striking in Table 1 is the magnitude of the standard deviation of the bias: 0.96 percentage points. This means that while, on average, the CPI inflation rate is biased upwards by 0.6 percentage points per year, one can only say with 95 percent confidence that this bias lies between -1.5 and 2.8 percentage points. In other words, if the official inflation

Figure 5



rate is one percent per year and aggregate CPI errors are the same as those for grocery items, one can only infer that the true is inflation rate is between -1.5 and 2.8 percent. Thus, a one percent measured inflation rate would not be sufficient information for a central bank to know if the economy is in inflation or deflation. The magnitude and standard deviation of the bias in the grocery component of the CPI is particularly striking since other components of the CPI are subject to much larger measurement errors, potentially yielding an even larger and more noisy bias in the overall CPI.[13]

We can see the magnitude of the noise in Figure 5. The figure plots the difference between the CPI and the Törnqvist index over the full time period. There are two important findings revealed by the figure. First, as we saw in Table 1, the CPI bias is not constant. While the bias is positive on average, the difference between the CPI and the Törnqvist fluctuates. Second, the magnitude of the measurement error is quite large in comparison with the

[13]Of course, averaging over more price series might improve the picture for the overall CPI—a point we will address later—but the magnitude of the errors is still worrying.

underlying Japanese inflation rate. While one can see that the bias is clearly positive on average, there appears to be no clear relationship between the magnitude of the bias and the underling inflation rate. For example, the bias was not exceptionally high or low in either of the two periods that Japan experience inflation peaks, 1990-1 and 2008-9. The plot suggests that it is not implausible to think of this noise as classical measurement error.

The results in Figure 5 and Table 1 have important implications for our understanding of inflation. As basic econometrics tells us, the fact that there is substantial measurement error in the CPI means those who assume that a one percent movement in the CPI corresponds to a one percent movement in inflation will systematically overstate true inflation movements. We turn to formalizing this intuition in the next section.

3 Econometric Theory

3.1 Inferring True Inflation from Measured Inflation

In general, policy makers do not observe bar-code data, so the relevant question is how do we infer true inflation from the measured values of inflation. We think about this problem mathematically by first assuming that measured inflation—the CPI, which we denote π_t^{CPI}—is equal to true inflation, which we denote π_t^T, plus some measurement error, ϕ_t:

$$\pi_t^{CPI} = \pi_t^T + \phi_t \tag{1}$$

where π_t^T and ϕ_t are both unobserved. Equation (1) can be thought of as the standard equation used in the CPI bias literature.[14] However, most users of the CPI are are interested in a related but different question: what is the expected value of true inflation given measured inflation, or $E\left[\pi_t^T|\pi_t^{CPI}\right] = \alpha' + \beta'\pi_t^{CPI}$? Since π_t^T and π_t^{CPI} are both univariate and they

[14] One can think of this formally by realizing that the price measurement literature has focused on obtaining a link between true and measured inflation, i.e. estimating $\pi_t^{CPI} = \pi_t^T + \alpha + \epsilon_t$, where π^{CPI} is the CPI inflation rate, π_t^T is the true inflation rate, α is the bias, and ϵ_t is a mean-zero error term. This approach is equivalent to equation (1) with $\phi_t = \alpha + \epsilon_t$.

follow a linear relationship, the minimum mean square error estimate of π_t^T is its conditional expectation, which we can express as:

$$E\left(\pi_t^T|\pi_t^{CPI}\right) = E\left(\pi_t^T\right) + \frac{Cov\left(\pi_t^T, \pi_t^{CPI}\right)}{Var\left(\pi_t^{CPI}\right)}\left[\pi_t^{CPI} - E\left(\pi_t^{CPI}\right)\right] \tag{2}$$

indicating that the extent to which an unexpected change in the CPI is a signal of an unexpected change in true inflation is governed by the ratio of the covariance of inflation and the CPI to the variance of the CPI.

The amount of information on true inflation, π_t^T, that is contained in observed inflation, π_t^{CPI}, therefore depends crucially on the covariance of the two series relative to the variance of the observed inflation series. It is useful to rewrite equation (2) in terms of a regression coefficient, β, that would be obtained from regressing π_t^T on π_t^{CPI}:

$$E\left(\pi_t^T|\pi_t^{CPI}\right) = \left[E\left(\pi_t^T\right) - \beta E\left(\pi_t^{CPI}\right)\right] + \beta\pi_t^{CPI}$$

where:

$$\beta \equiv \frac{Cov\left(\pi_t^T, \pi_t^{CPI}\right)}{Var\left(\pi_t^{CPI}\right)} = \frac{Var\left(\pi_t^T\right) + Cov\left(\pi_t^T, \phi_t\right)}{Var\left(\pi_t^T\right) + Var\left(\phi_t\right) + 2Cov\left(\pi_t^T, \phi_t\right)} \tag{3}$$

where the second equality follows from equation (1).[15]

Equation (3) forms the foundation of everything that follows. In particular, the notion that a given movement in the CPI produces a movement in actual inflation of the same magnitude, i.e. $\beta = 1$, relies on the assumption that there is no variation in the measurement error of CPI. If there is *any* volatility measurement error in the CPI, true inflation cannot be expressed as equal to measured inflation less some constant term. In other words, the fact that the bias term in Figure 5 is not a horizontal line, provides *prima facie* evidence that the

[15] This is analogous to the signal extraction problem studied in Lucas [1972] for a case in which a firm must learn about the aggregate price level from the price of a good the firm produces. Lucas [1973] shows that the signal-to-noise ratio, which is defined as the ratio between the variance of aggregate shocks and the variance of idiosyncratic shocks, differs across countries and tends to be higher for countries with a higher variability of aggregate demand..

common practice of assuming true inflation equals measured inflation less some bias term is wrong.

The noisy indicators literature has examined cases where β is constant. We will next show that one should not expect this. In order to do this, it will be useful to build some theory from first principles on what we should expect the relationship to be.

3.2 Decomposing "True" Inflation

We need to make a few assumptions about the underlying error structure causing CPI mismeasurement in order to understand why β in equation (3) is unlikely to equal one or be constant. We begin by defining π_{it} as the Törnqvist index inflation rate for a *lower-level* index for category i (*e.g.,* milk, cheese, etc.); this index is computed by aggregating many bar codes. Aggregate inflation, π_t^T, is given by the following index:

$$\ln\left(1+\pi_t^T\right)=\sum_{i=1}^{n} w_{it}\pi_{it} \tag{4}$$

where n denotes the number of goods categories in the *upper-level* of the price index; s_{it} is the expenditure share of good i in period t; the Törnqvist weight is then given by $w_{it}=\frac{1}{2}\left(s_{it}+s_{it-1}\right)$; and $1+\pi_{it}$ is the relative price of good i in period t relative to its price in period $t-1$, and $\sum_{i=1}^{n} w_{it}=1$. The Törnqvist price index is exact for the translog expenditure system, which is a second order approximation to an arbitrary, twice-differentiable, homothetic demand system. We therefore impose this structure on the data and write the expenditure share of good i in period t as

$$s_{it}=\alpha_i+\gamma\left(\ln p_{it}-\overline{\ln p_t}\right) \tag{5}$$

where α_i is a parameter that is related to the quality of good i, γ is a parameter related to the demand elasticity, p_{it} is the price of the good, and $\overline{\ln p_t}$ is the average log price of all goods in the economy. The price change of every good i from period t to period $t-1$ can be decomposed as follows:

$$\pi_{it} = \mu_t + \nu_{it}, \tag{6}$$

where μ_t is the aggregate component of inflation and ν_{it} is an idiosyncratic component of that good's price change. This enables us to rewrite the Törnqvist inflation index as

$$\pi_t^T \;\; = \;\; \mu_t + \sum_{i=1}^{n} w_{it}\nu_{it}. \tag{7}$$

3.3 The Variance of "Official" Inflation

As we have argued earlier, statistical agencies cannot compute the Törnqvist index in equation (4) because they do not collect the quantity data needed to compute π_{it} and the weights they use do not conform exactly to w_{it}. We can think of official inflation rates as weighted averages of price changes in which the weights and formulas contain errors. We can therefore write a general equation for the CPI, π_t^{CPI} in terms of the true underlying price indexes as:

$$\pi_t^{CPI} = \sum_{i=1}^{n} \left(w_{it} + \epsilon_{it}\right)\left(\pi_{it} + \delta_{it}\right), \tag{8}$$

where we use ϵ_{it} to denote the error in the upper-level weight for good i at time t and δ_{it} to denote the lower-level error in the measured inflation rate for the good. In this formulation, if the the upper- and lower-level errors were eliminated, the CPI inflation rate would collapse to that of the Törnqvist.

In order to make the analysis that follows tractable, we need to assume that these errors are random independent draws from distributions with zero means. In particular, we assume $E\left(\epsilon_{it}\right) = E\left(\delta_{it}\right) = 0$, $Cov\left(\epsilon_{it}, \epsilon_{jt}\right) = Cov\left(\delta_{it}, \delta_{jt}\right) = Cov\left(\nu_{it}, \nu_{jt}\right) = 0$, and $\mu_t \perp \epsilon_{it}, \delta_{it}$. As we show in Appendix A, these assumptions are sufficient to ensure that the error term in equation (3) is independent of the level of inflation, a condition that we will examine in the empirical section. These assumptions enable us to derive the variance of the the CPI in terms of the signal, given by variance of true inflation ($\sigma^2_{\pi^T_t} \equiv var\left(\pi^T_t\right)$) and the various errors given below:

$$
\begin{aligned}
Var(\pi^{CPI}_t) \;&=\; \sigma^2_{\pi^T_t} + \sigma^2_{\delta_t}\left(\sum_{i=1}^{n} s^2_{i,t-1} + \frac{\gamma^2}{4}\left(n-1\right)\sigma^2_{\nu_t}\right) \\
&\quad + n\sigma^2_{\epsilon_t}\left[\sigma^2_{\mu_t} + \sigma^2_{\nu_t} + \sigma^2_{\delta_t}\right], \qquad (9)
\end{aligned}
$$

where the algebra used to derive this equation is relegated to Appendix Appendix A.

Equation (9) is key to understanding the problems of formula bias in computing the variance of standard measures of inflation. The first surprising feature of equation (9) is that number of price quotes does not appear directly on the right-hand side. Implicitly, however, the number of price quotes could affect the variance in the lower-level price index, $\sigma^2_{\delta_t}$. However, even in infinite samples, this variance will not in general be zero because δ_{it} is the difference between measuring inflation with Törnqvist index and the actual index used by the statistical agency. For example, the U.S. PCE deflator is based on a geometric average of prices, which is the same price index that arises if one assumes that utility is Cobb-Douglas. In general, a consumer evaluating a set of price changes with Cobb-Douglas preferences will report a different inflation rate than one with translog preferences unless the elasticity of substitution for all goods equals one. The extent of the difference is an empirical question that is the focus of this paper, but theoretical difference is clear. The second interesting feature of equation (9) is that as long as the variance of the upper-level and lower-level CPI errors are not jointly zero (*i.e.,* $\sigma^2_{\delta_t} = \sigma^2_{\epsilon_t} = 0$), the variance in measured CPI will exceed that of the true inflation rate. This fact is the crucial element that will break the one-to-one link between movements in measured and actual inflation.

In order to show this, we need to compute correlation between π^{CPI}_t and π^T_t. In Appendix Appendix A, we show that under reasonable assumptions about the independence of the various idiosyncratic shocks we can write:

$$Cov\left(\pi_t^T, \pi_t^{CPI}\right) \;=\; \sigma^2_{\pi_t^T} \tag{10}$$

The key assumption behind this result is that the CPI measurement errors, ϵ_{it} and δ_{it}, are independent draws from a distribution and, therefore, uncorrelated with the components of true inflation, μ_t and ν_{it}. While equation (10) does not need to be true for all possible price shocks, we will show in the empirical section that it is an extremely good approximation of reality. This result has important implications for our understanding of equation (2). In particular, if we maintain the assumptions necessary to derive equation (10), we can rewrite the expression for β derived in equation (3) in terms of the fundamental microeconomic price shocks as,

$$\begin{aligned}
\beta &= \frac{Var\left(\pi_t^T\right)}{Var\left(\pi_t^T\right) + Var\left(\phi_t\right)} \\
&= \frac{\sigma^2_{\pi_t^T}}{\sigma^2_{\pi_t^T} + \underbrace{\sigma^2_{\delta_t}\left(\sum_{i=1}^{n} s^2_{i,t-1} + \frac{\gamma^2}{4}\left(n-1\right)\sigma^2_{\nu_t}\right) + n\sigma^2_{\epsilon_t}\left[\sigma^2_{\mu_t} + \sigma^2_{\nu_t} + \sigma^2_{\delta_t}\right]}_{\text{“CPI Noise”}\geq 0}} \leq 1
\end{aligned} \tag{11}$$

where $\sigma^2_{\mu_t}$ is the variance in common price shocks inflation, and $\sigma^2_{\nu_t}$ is the variance in idiosyncratic price shocks.

Equation (11) provides a mathematical formula for understanding how common and idiosyncratic price shocks as well as lower- and upper-level price index errors affect the relationship between measured and true inflation. A second implication is that with non-constant measurement errors (*i.e.*, $\sigma^2_{\delta_t} > 0$ or $\sigma^2_{\epsilon_t} > 0$) changes in the CPI always overestimate actual inflation changes.

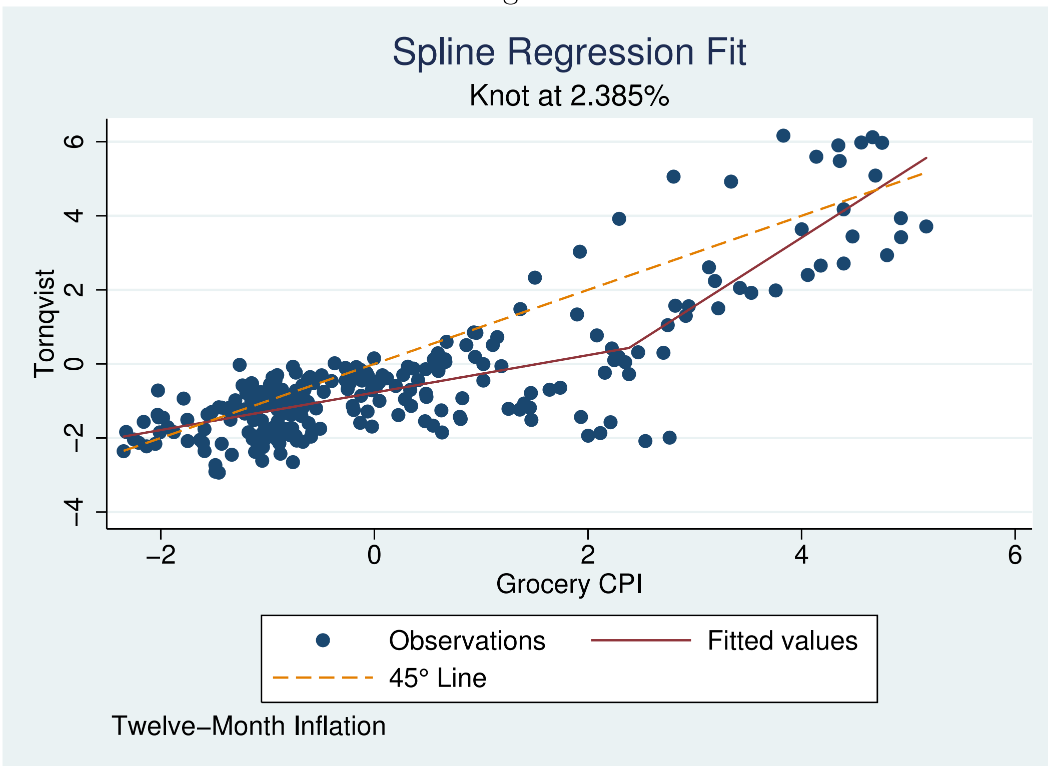
Our ability to decompose this bias into the underlying components enables us to make much stronger statements about the relationship between true inflation and measured inflation. Ball et al. [1988], among others, have argued that the variance of inflation should rise as inflation rises. However, they were only able to make this claim using U.S. CPI data, and thus, it is not clear that this rise in variance is due to actual inflation becoming more volatile, i.e., a rise in $\sigma^2_{\mu_t}$, or just an increase in the volatility of idiosyncratic price movements, $\sigma^2_{\nu_t}$. The former might occur if monetary shocks become more volatile when inflation rises, and the latter would arise if price dispersion rises with inflation. One of the interesting elements of our setup is that we can not only examine if $Cov\left(\sigma^2_{\pi^T_t}, \pi^T_t\right) > 0$, but we can also understand why it is true.

A second, and more important, feature of our approach is that we also have a direct measure of the "CPI Noise" term in equation (11). If $Cov\left(\sigma^2_{\pi^T_t}, \pi^T_t\right) > 0$ and this term is approximately constant, then this implies that there will be a non-linear relationship between CPI inflation changes and actual inflation changes. When the CPI is low, the signal-to-noise ratio will be low because most movements in the CPI will be driven by noise, and one should not adjust expectations of true inflation from movements in the CPI. But when inflation is high and $Var\left(\pi^T_t\right)$ is large, movements in our estimate of actual inflation much more with movements in the CPI. In order to compute the exact magnitudes we need to first compute the elements of equation (11).

4 Results

We divide our results into several sections. First, we document that there is a strong non-linear relationship between true inflation and the CPI. Second, we investigate the microstructure of this non-linearity with an aim of understanding what elements of measured inflation explain our findings. Third, we conduct a series of robustness checks to examine whether the results stem from the methodologies employed by statistical agencies or whether our results simply stem from using a different dataset to measure prices. Finally, we consider some potential fixes to the measurement of inflation.

Figure 6



4.1 Inferring Inflation

We now turn to studying the relationship between the official CPI and the Törnqvist index addressing a series of questions motivated by the theoretical framework outlined above:

1. Is the signal-to-noise coefficient (β) equal to one?
2. Does true inflation move less than one-for-one with measured inflation when measured inflation (and its variance) is low?
3. Does this problem disappear when inflation is high?

Fortunately, all of these questions can be qualitatively assessed by looking at the data. We start by plotting π_t^T against π_t^{CPI} in Figure 6. Each point in the plot represents a 12-month inflation rate taken from our sample. There is a very strong positive relationship between true inflation and CPI inflation when the inflation rate exceeds 2 percent, but there is a much weaker connection between CPI inflation and actual inflation when the CPI is registering rates of inflation below 2 percent per year.

Table 2
Tornqvist vs. CPI Inflation

	Tornqvist	Tornqvist	Tornqvist	Tornqvist
Grocery CPI	0.833***	0.553***		
	(0.148)	(0.151)		
Grocery CPI^2		0.119**		
		(0.0494)		
Grocery CPI ($\leq$ 1%)			0.398**	
			(0.189)	
Grocery CPI ($>$ 1%)			1.257***	
			(0.241)	
Grocery CPI ($\leq$ 2.385%)				0.506***
				(0.165)
Grocery CPI ($>$ 2.385%)				1.844***
				(0.445)
Constant	-0.584***	-0.907***	-0.893***	-0.774***
	(0.173)	(0.224)	(0.216)	(0.183)
Observations	262	262	262	262
Adjusted R^2	0.668	0.718	0.714	0.739

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Regression evidence confirms that our eyes are not deceiving us. In Table 2 we present a number of regressions of of π_t^T against π_t^{CPI}. In the first column we perform a simple OLS regression and obtain a coefficient of 0.83. The fact that this coefficient is less than one suggests that CPI measurement error could be causing an attenuation bias. However, simple inspection of Figure 6 indicates that this bias is not stable. In column 2 we regress π_t^T against π_t^{CPI} and its square. The data indicate that the attenuation bias is much larger when CPI inflation is low relative to when it is higher. In column 3, we run a spline regression with a knot placed at 1 percent measured inflation. The interesting feature of this regression is that for inflation rates below 1 percent, there is a much weaker relationship between measured and actual inflation. If we chose an endogenous knot based on maximum likelihood estimation, the data place it at 2.4 percent, indicating that there is a substantially weaker relationship between measured inflation and true inflation below this point.[16]

The results presented in the last column suggest that the relationship between measured and actual inflation is highly non-linear. Consider, for example, how to interpret a move in the Japanese CPI from -1 to 2 percent. According to the regression evidence in the last column in Table 2, an inflation rate of -1 percent per year—as the Japanese CPI averaged for many years—corresponds to a true inflation rate of -1.3. In other words the bias in inflation is very small when inflation is moderately negative. However, if CPI inflation were to rise by three percentage points to 2 percent, the negative bias would rise in magnitude dramatically, reaching 1.8 percentage points. Our point estimate suggests that two percent measured inflation rate is only a 0.2 percent inflation rate. *In other words the current Bank of Japan target inflation rate of 2 percent is extremely close to the correct rate to achieve price stability!* Moreover, the fact that the bias point estimate varies considerably with the underlying inflation rate means that one should exercise caution in comparing estimates of the bias computed using samples composed of observations using different inflation rates.

A second striking feature of the non-linearity that we observe is that further increases in CPI inflation imply much sharper rises in true inflation. For example, our estimation implies that an increase in inflation from 2 percent to 5 percent would correspond to an increase in true inflation from 0.24 percent to 3.4 percent. Since the upward bias largely disappears at higher inflation rates, central banks should pay much more attention to inflationary changes when inflation is high than when it is close to zero. A central bank that deems a movement in CPI inflation from 0 to 2 percent as the same as a movement from 2 to 4 percent is liable to dramatically overreact to inflation when it is low and and under react when it is high.

[16] It is possible that the relationship between measured inflation and true inflation depends on the absolute magnitude of inflation, rather than its level. Suppose, for example, that inflation volatility is increasing in the magnitude of inflation, rather than its level, that is, $Cov\left(Var\left(\pi_t^T\right), |\pi_t^T|\right) > 0$ rather than $Cov\left(Var\left(\pi_t^T\right), \pi_t^T\right) > 0$. We considered a specification with two knots symmetrically placed around 0, but the data did not support this parameterization instead selecting a model with a single knot placed at 2.4 percent. This may reflect the fact that we do not experience any highly deflationary periods.

4.2 The Microstructure of Inflation Measurement Errors

The analysis above gives visual and econometric evidence for a non-linear relationship between the CPI and actual inflation, but we still have not identified the micro-foundations of this relationship. Our theory depends on the CPI measurement error being classical. Theoretically, one can construct cases in which this is not true, but its veracity is an empirical question. In order to assess whether $Cov\left(\pi_t^T, \phi_t\right) = 0$, we regressed ϕ_t on π_t^T and could not statistically reject that there was no relationship between the measurement error and the level of inflation. Since we cannot statistically reject the hypothesis that the inflation measurement error is classical, we assume it is so and proceed with our decomposition.

Table 3
Summary Statistics

	All	Ratio (1% Knot)	Ratio (2% Knot)	Ratio (Endogenous Knot)
$\beta(\pi^{\text{CPI}}, \pi^T)$	0.833	2.56	3.42	3.65
$\sigma^2_{\pi_t^T}$	3.18e-04	10.41	8.01	5.68
$\sigma^2_{\nu_t}$	3.49	2.46	2.55	2.34
$\sigma^2_{\hat{\mu}_t}$	0.79	7.45	5.77	4.16
$\sigma^2_{\varepsilon_t}$	0.32	1.14	1.19	1.35
$\sigma^2_{\delta_t}$	2.83	1.57	1.60	1.62
Herf_t	0.019	1.05	1.05	1.05
CPI Noise	0.20	2.36	2.49	2.56
$\#\text{Items}_t$	135.2	0.97	0.98	0.99
γ_t	-0.68			
Observations	262	0.32	0.21	0.15

Note: Entries for $\sigma^2_{\nu_{i,t}}$, $\sigma^2_{\hat{\mu}_t}$, $\sigma^2_{\varepsilon_{i,t}}$, $\sigma^2_{\delta_{i,t}}$, and the CPI Noise (defined in Equation 12) are divided by the entry for $\sigma^2_{\pi_t^T}$. The "Ratio" columns give the ratio of a variable in periods where π^{CPI} is greater than the knot value to that variable's value calculated from observations in those months when π^{CPI} is less than the knot. The "Endogenous Knot" was determined to be 2.385%.

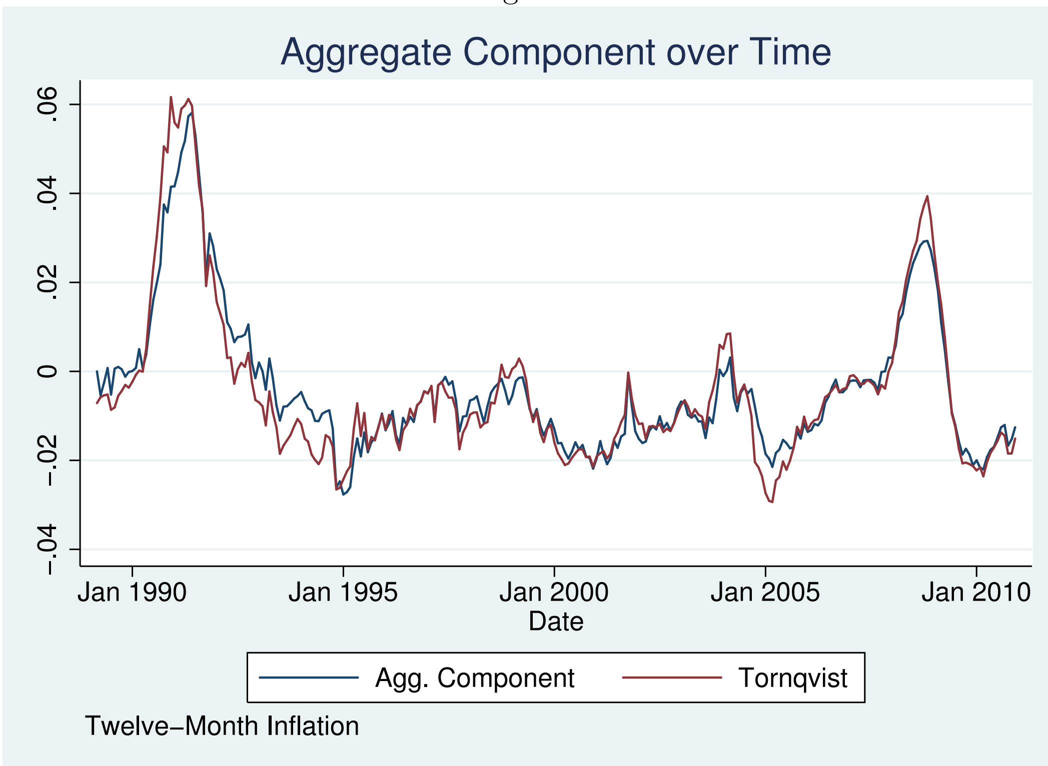
Equation (11) provides the formula for decomposing β into the underlying variances. Fortunately, our dataset is rich enough that we can directly measure all of the variance components that are likely to create the biases. In Table 3 we compute $\hat{\beta}$ along with all of its components to examine why it varies with the level of inflation. In the first column of the table, we present the results from the full sample with the first three elements of the column presenting our regression coefficient from the first column of Table 2 along with its decomposition as given in equation (11).

A critical feature of equation (11) is that it tells us how to decompose movements in the link between actual and CPI inflation based on movements in the underlying forces driving movements in variance of actual inflation ($\sigma^2_{\pi^T_t}$) and movements in the CPI noise term. Let's begin by focusing on the determinants of inflation volatility.

We know from equation (6) that we can decompose actual inflation of an item, π^T_{it}, into two components: the aggregate component of inflation, μ_t, and its idiosyncratic component, ν_{it}. By simply running this OLS regression we can identify each element of this decomposition. In Figure 7, we plot the inflation rate implied by the aggregate component of inflation, μ_t, against that of true inflation, π^T_t. Not surprisingly, these two series track each other extremely closely, which is what one would expect if aggregate inflation were driven largely by shocks common to all sectors rather than to shocks to a few sectors.

In the next cells of the first column of Table 3, we express each of the component variances as a share of $\sigma^2_{\pi^T_t}$, so that one can get some sense of how important each factor is relative to the variance of true inflation. The fact that the variance of the aggregate component of inflation, $\sigma^2_{\mu_t}$, is 79 percent as large as $\sigma^2_{\pi^T}$ implies that actual inflation movements are an important determinant of CPI inflation movements as one might suspect. It is a little harder to interpret the importance of idiosyncratic inflation volatility because it does not enter linearly, but the fact that we saw in Figure 7 that π^T_t tracks μ_t closely means that these idiosyncratic shocks to individual inflation rates cannot be an important determinant of actual inflation. The next two cells show the importance of error terms that only affect the variance of measured (CPI) inflation but not actual inflation. As one can see from the table, the variance of the lower-level errors, $\sigma^2_{\delta_t}$, are much larger than the upper-level ones, $\sigma^2_{\epsilon_t}$, although here, too, it is difficult to assess the importance of each error because they affect the CPI noise interactively.

Figure 7



In order to make some progress on this issue, we can think of performing a number of counterfactual exercises in which we assume that a statistical agency could eliminate different types of errors and see what this does to the variance of the aggregate measurement error. For example, if a statistical agency eliminated the upper-level weighting problems, it would set $\sigma^2_{\epsilon_t} = 0$ in equation (11), we can compute how much the variance in the CPI noise would fall using the correct upper-level weights. Our results suggest that a perfect weighting structure would only reduce the variance in CPI noise by 49 percent. However, if we were to eliminate lower-level measurement errors, *i.e.* setting $\sigma^2_{\delta_t} = 0$, this would reduce the variance in CPI noise by 71 percent.[17] These calculations make clear that at least half of the problem in the CPI is that there are substantial formula biases and other measurement errors at the lower level. The importance of lower-level errors in driving overall CPI measurement error probably stems from the fact that the inability to weight goods by sales and the choice of

[17]The two numbers do not sum to 100 percent because the $n\sigma^2_{\epsilon_t}\sigma^2_{\delta_t}$ term can be eliminated by setting either $\sigma^2_{\epsilon_t}$ or $\sigma^2_{\delta_t}$ to zero.

the price aggregation method at the lower level is not rigorously based on theory.

In addition to decomposing the sources of the errors, we also can back out the combined impact of each of these terms by using equation (11) and our parameter estimates to solve for the variance in the CPI noise. This number is presented in the row labeled "CPI Noise" in Table 3 and indicates that this noise is about 20 percent as big as the variance of actual inflation implying a signal-to-noise ratio of about 5.[18]

The following three columns of Table 3 present what happens to each parameter as we divide the data into different bins. The first two columns corresponds to results of assuming that there are knots at inflation rates of one percent and two percent, and the last column corresponds to the endogenous knot we identified in Table 2. For example, we see in the first row that if we estimated β using data in which the CPI exceeded 2.4 percent (the value of our endogenous knot) we would obtain a value that is 3.65 times larger than if we estimated β using data in which measured inflation is less than 2.4 percent. The lower rows provide information how changes in the key variances that determine β in equation (11) vary above and below each knot. The second row of the table shows that the variance of true inflation, $\sigma^2_{\pi^T}$, rises by a factor of 5.7 to 10.4 (depending on the cutoff) as we move from a low inflation regime to a high inflation regime. This positive association between inflation and inflation volatility is the standard result that has been documented elsewhere.

However, we are able to drill deeper into this result than past studies and understand its cause. As one can see from the third and fourth rows of Table 3 the result is driven by two forces. First and foremost, we see that the variance of the aggregate component of inflation, $\sigma^2_{\mu_t}$, rises by a factor of 4 in the endogenous knot specification, which is consistent with the idea that the variance of aggregate shocks rise as inflation rates rise. Second, we see that the variance of idiosyncratic price shocks, $\sigma^2_{\nu_t}$, doubles, which suggests that rising price *dispersion*—a feature of many staggered pricing models—also increases with inflation.

[18] As a check on our results we can also solve for the translog demand parameter, γ, which comes out equal to -0.68, implying that firm with a 20 percent market share would have a 29 percent markup. Our procedure is surely not the most efficient means of estimating this demand parameter; we report the number simply to demonstrate that our approach does not imply an implausible elasticity.

While the data also indicate that the variance of CPI noise rises with inflation, the increase is much smaller: as we move from low- to high-inflation regimes, CPI noise only rises half as much as the rise in variance of true inflation. This increase comes from a number of sources. The rise in idiosyncratic price shocks, $\sigma^2_{\nu_t}$, interacts with the substantial rise in lower level measurement errors, $\sigma^2_{\delta_t}$, reinforcing the level of measurement error. Similarly, upper level weighting errors, $\sigma^2_{\epsilon_t}$, also rise slightly, further boosting measurement error. However, despite these effects, the CPI's signal-to-noise ratio almost doubles in the high inflation regime, which explains why the CPI is a much more reliable predictor of inflation when inflation is high.

5 Robustness

Thus far, we have been focused on understanding the biases and measurement error in the Japanese CPI. However, this raises questions of how general these results are. In this section, consider a number of possible explanations and improvements that might mitigate the problems that we have identified. In Section 5.1, we consider the extent to which the errors we have identified are due to the use of two different datasets (Nikkei POS and the SBJ sample), rather than pure methodological differences. We confirm that using the SBJ sampling and index construction method using Nikkei POS data produces qualitatively the same results, which leads us to conclude that the key driver of the results is the official index construction methodology. In Section 5.2, we sharpen this result by conducting a counterfactual exercise in which we examine how the results might have differed if the Japanese government implemented the U.S. methodology for computing inflation in the PCE. Since geometric averaging and random sampling should reduce the formula and sampling errors, we expect a partial improvement. Indeed, we find that implementing mitigates but does not eliminate the problem. Finally, we check to make sure our results are not driven by an anomoly in Japanese bar-code data. In Section 5.3, we replicate these results using U.S. data to see if we observe similar patterns in U.S. bar-code data.

5.1 Data Comparison

An obvious concern about our results is that we are working with two different datasets—the Nikkei POS and the CPI—and that our results may be due to the fact that the data collection methods differ. In order to test whether this is driving our results we need to replicate the Japanese CPI using the Nikkei POS data from 1988 to 2010. This is not a trivial exercise as the purposive sampling method used by SBJ uses a non-random selection of goods. Fortunately, Imai et al. [2012] used the CPI price quote descriptions to identify the bar codes that match these specifications in the Nikkei data and then replicated the Japanese CPI methodology on our data.[19] Despite our best efforts to replicate the methodology used by the SBJ to select goods and stores, our method still differs from the official one because of the many judgement calls used in the SBJ sample selection procedure that are not clearly documented. Although the errors are different, our replicated CPI is as closely correlated with the Törnqvist as the official index. The adjusted R^2 arising from regressing the Törnqvist on the CPI barely changes: rising from 0.67 in the first column of Table 2 to 0.69 in Table 4. The similarity suggests that most of the differences between the two series are likely to be attributable to formula differences.

Table 4 reports the results of replicating the regressions in Table 2 on the replicated Japanese CPI. A striking feature of these results is how similar the two sets of results are. One cannot statistically reject that the coefficients in in Table 4 are the same as their counterparts in in Table 2. These results indicate that the reason for the non-linearity is not based on differences in the underlying price data in the Nikkei and CPI samples, but must emanate from the different methodologies used to construct the price indexes.

[19] The replicated index is constructed with the purposive sample of jancodes, using prices drawn from a sample of outlets that is "re-set" each time that the base weights are re-set monthly data (for each item, we will take the 42 outlets with the highest sales of purposively-sampled jancodes for the item in the year before the re-basing). We use monthly unit values as prices. A Dutot index is used to aggregate the prices to the item-month level and official weights are used to aggregate across items.

Table 4
Replication Using Only Bar-Code Data: Tornqvist vs. Japanese CPI Methodology

	Tornqvist	Tornqvist	Tornqvist	Tornqvist
Replicated CPI	0.730***	0.605***		
	(0.116)	(0.0610)		
Replicated CPI^2		0.0850***		
		(0.0223)		
Replicated CPI ($\leq$ 1%)			0.380***	
			(0.0829)	
Replicated CPI ($>$ 1%)			1.208***	
			(0.176)	
Replicated CPI ($\leq$ .745%)				0.361***
				(0.0816)
Replicated CPI ($>$ .745%)				1.160***
				(0.165)
Constant	-0.256	-0.617***	-0.655***	-0.685***
	(0.172)	(0.171)	(0.178)	(0.178)
Observations	228	228	228	228
Adjusted R^2	0.664	0.733	0.739	0.739

11-lag Newey-West standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.2 From Dutot to Jevons: Does Geometric Averaging Improve Measurement?

There have been a number of analyses that have suggested that one way to improve the CPI is to move from Dutot index of prices (which is comprised of a simple average of prices) to a Jevons index (which uses a geometric average of prices) at the lower level. Both methods are acceptable according to ILO standards, and the ILO reports that slightly more countries use the Dutot method than the Jevons. For example, in addition to Japan, Belgium, Germany, and the U.K. also use Dutot indexes. However, many countries, including the U.S., have recently switched to the Jevons methodology because it better controls for consumer substitution. It therefore is reasonable to ask whether a methodology akin to that used by the U.S. in the construction of the PCE deflator eliminates the non-linearity. In order to assess the importance of arithmetic versus geometric averaging, we replicated the methodology used by the BLS in the selection of the price quotes, and then constructed a price index based on the PCE deflator that consisted of geometric averaging at the lower level and a Törnqvist index at the upper level.[20]

The results from this exercise are presented in Table 5. If we compare our results based on U.S. methodology with those using Japanese methodology in Tables 2 or 5, we see that using geometric averaging improves our measure of inflation both in terms of raising the R^2 and in terms of reducing the level of the bias. The point estimates from Table 2 imply that if the inflation rate using the Japanese methodology is one percent, the upward bias is 1.1

[20] In order to do this we replicated both the sample size and the sampling procedure used by the BLS. The BLS collects prices for 85,000 products each month. These products are classified into 305 entry level items (ELIs), of which 62 are food as opposed to 135 items in Japan. We chose the number of store-bar-codes according to the following formula:

$$\frac{\text{62 Food ELIs}}{\text{305 Total ELIs}} \times \frac{\text{85,000 price quotes}}{\text{135 Japan CPI items}} = \text{128 store-barcodes per Japan CPI item}$$

The BLS chooses products for inclusion on a rotating basis. The BLS methodology replaces 1/16th of their sample of price quotes each quarter, and we replicated this in our data. We did this by resampling 8 store-bar-codes in each of our item categories each quarter. Selection probability was a store-bar-code's share of the previous quarter's sales within its CPI item. Sampling weights were based on the sales of each bar code in the two years prior to each base year.

Table 5
Replication Using Only Bar-Code Data: Tornqvist vs. US PCE Methodology

	Tornqvist	Tornqvist	Tornqvist	Tornqvist
Replicated PCE	1.034***	0.925***		
	(0.0676)	(0.0257)		
Replicated PCE^2		0.0700***		
		(0.00760)		
Replicated PCE ($\leq$ 1%)			0.783***	
			(0.0568)	
Replicated PCE ($>$ 1%)			1.370***	
			(0.0530)	
Replicated PCE ($\leq$ 1.752%)				0.819***
				(0.0499)
Replicated PCE ($>$ 1.752%)				1.542***
				(0.0550)
Constant	-0.222***	-0.441***	-0.464***	-0.419***
	(0.0728)	(0.0680)	(0.0760)	(0.0659)
Observations	240	240	240	240
Adjusted R^2	0.947	0.967	0.965	0.967

11-lag Newey-West standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

percent, but the upward bias is only 0.6 percent using the U.S. PCE methodology based on the estimates in Table 5. Moreover, when inflation is low, true inflation only rises 36 percent as fast as measured when using the Dutot measure of inflation but 82 percent as fast when it is measured using the U.S. procedure (as one can see from the last columns of Tables 4 and 5).

One obvious question is how important are sampling errors versus formula errors in the generation of measurement errors in price indexes. We can easily address this question in the case of the PCE deflator methodology by maintaining the BLS sampling technique but switching the price aggregation method from a simple geometric average (the Jevons index) to the Törnqvist. By constructing this "PCE Törnqvist" index, we eliminate all of the formula bias in the price estimate but keep errors due to the BLS method of sampling a subset of prices. Table 6 reports the results of this exercise. Not surprisingly, eliminating the main source of formula bias makes the link between actual and measured inflation more linear, but it does not totally eliminate the curvature or the bias. Our results suggest that most of the problem in CPIs based on Dutot averages (as is used in Japan) is due to not allowing for substitution.

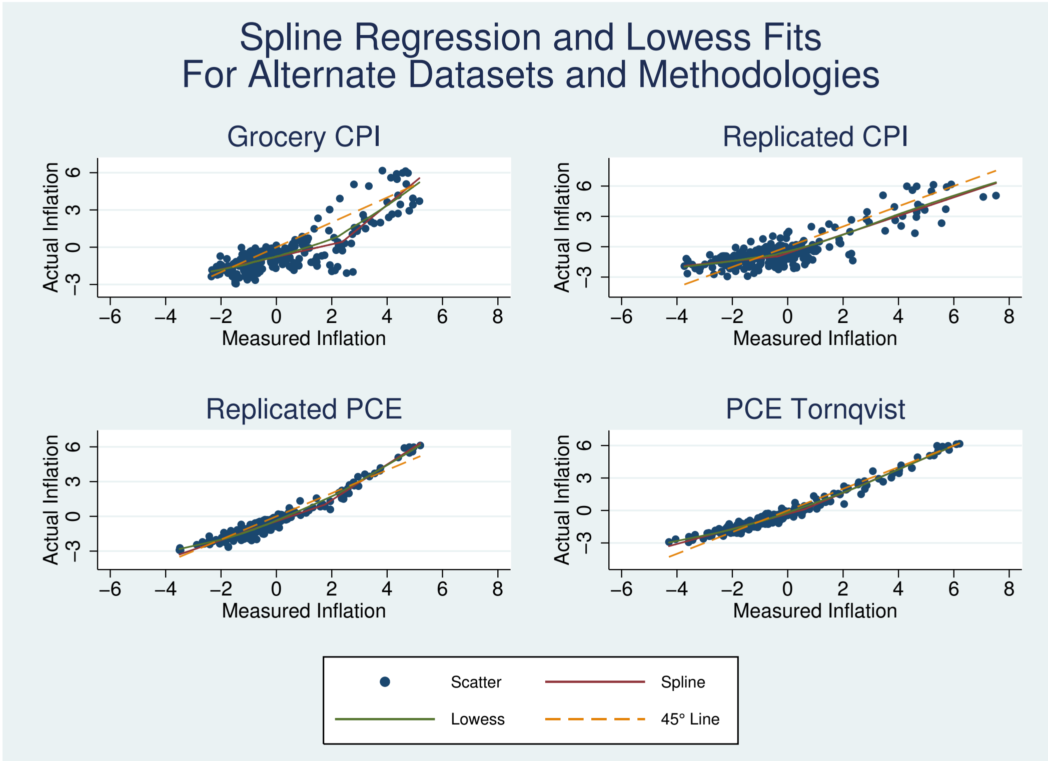
We can see this most clearly in Figure 8, which plots actual vs. measured inflation for our four ways of measuring inflation: Japan's actual grocery CPI, our replication of Japan's CPI, our replication of US methodology (using Nikkei data), and our PCE Törnqvist. In all plots, we observe a J-shaped data plot, indicating that in low inflation regimes, the link between measured and actual inflation is less than in high inflation regimes. However, as we move from using Dutot indexes in the top row, to a Jevons index in the lower left (the Replicated PCE), and finally to a Törnqvist in the lower right, the curvature lessens. The bottom line is that the relationship between true and measured inflation depends quite substantially on the measure used and these measures vary substantially across countries.

Table 6
Replication Using Only Bar-Code Data: Tornqvist vs. Tornqvist with US Sampling Methodology

	Tornqvist	Tornqvist	Tornqvist	Tornqvist
Tornqvist with US Sampling	0.888***	0.822***		
	(0.0514)	(0.0225)		
Tornqvist with US Sampling^2		0.0427***		
		(0.00435)		
Tornqvist with US Sampling ($\leq 1\%$)			0.712***	
			(0.0373)	
Tornqvist with US Sampling ($> 1\%$)			1.129***	
			(0.0369)	
Tornqvist with US Sampling ($\leq .527\%$)				0.692***
				(0.0343)
Tornqvist with US Sampling ($> .527\%$)				1.087***
				(0.0378)
Constant	-0.0602	-0.259***	-0.286***	-0.318***
	(0.0705)	(0.0559)	(0.0635)	(0.0604)
Observations	240	240	240	240
Adjusted R^2	0.952	0.971	0.970	0.970

11-lag Newey-West standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 8



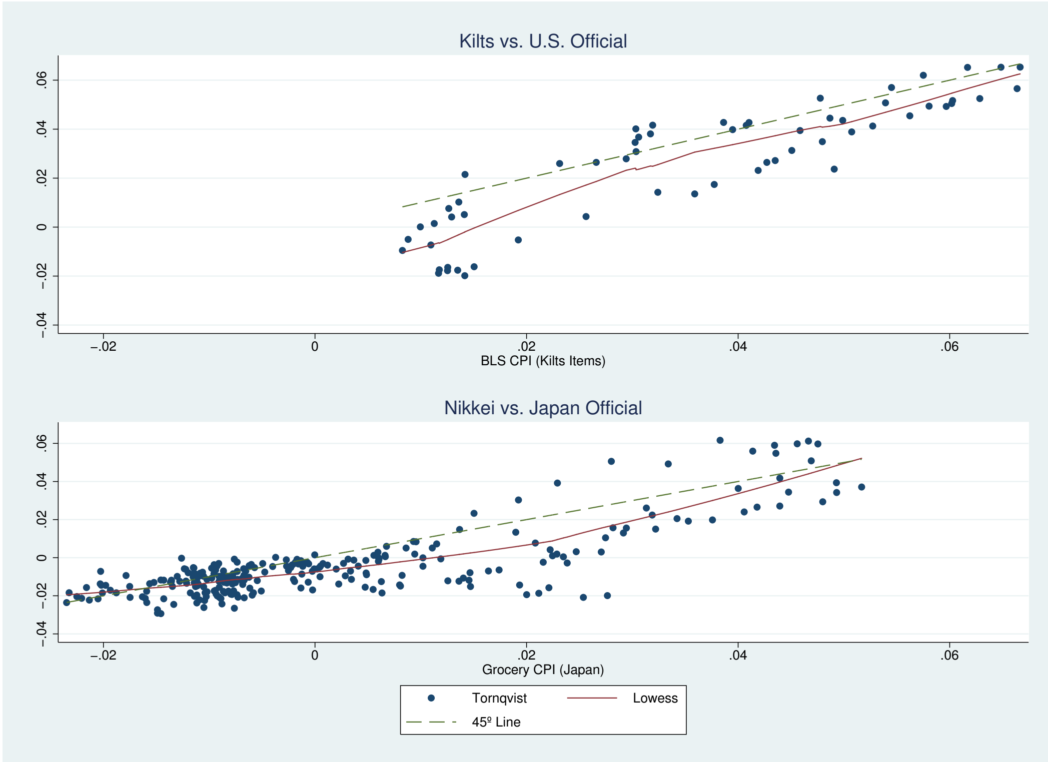
5.3 Robustness to U.S. Data

A final concern that one might have is that there may be a unique feature of Japanese data that is driving our results. In order to address this possible concern, we replicated our results using U.S. bar-code data drawn from Nielsen.[21] Unfortunately, the time series dimension of U.S. data is only seven years (as opposed to 23 years for Japan) and the U.S. lacks deflationary observations. These data limitations make it difficult to do careful statistical analysis like that on the Japanese data. However, we can present some plots to show that it appears that similar forces are at play in U.S. data.

Figure 9 presents the results of the Törnqvist inflation rate computed on U.S. Nielsen bar-code data against the U.S. CPI values for those items aggregated using the U.S. weights in the top panel and Figure 6 below for comparison. Both the U.S. and Japanese data indicate that, while there is very little bias in measured inflation above four or five percent,

[21]Details on the data can be found here: http://research.chicagobooth.edu/nielsen/datasets/consumer-panel-data.aspx

Figure 9



the upward bias in the U.S. CPI is around 1-2 percent when inflation is low. In the U.S. as in Japan, we see very little relation between measured and actual inflation when inflation is less than 2 percent. Moreover, this plot is quite consistent with what we observed in Table 6, which presented the results from estimating the bias in the U.S. methodology when implemented on Japanese data. In that table, we estimated that true inflation would be zero when the U.S. CPI methodology recorded an inflation rate of 1 percent, and in Figure 9 we find that when inflation in the U.S. CPI clusters around one percent the inflation rate in the Nielson sample is zero or negative. Thus, our results are robust to the usage of data from other countries.

6 Possible Solutions

Sections 6.1 and 6.2 consider two possible fixes for the problem. The first considers whether averaging over more categories of goods can eliminate the problem. The second considers whether computing inflation over longer time periods reduces the measurement problem.

6.1 Applicability of our results for the broader CPI

Another way of seeing that formula errors are driving our results more than sampling errors can be garnered from realizing that the CPI errors are not mitigated much by building indexes that use more categories of goods. We can obtain some sense of how much working with data for a larger set of items might matter by bootstrapping the underlying errors in the CPI and seeing how the variance of the CPI falls with as we average over more product categories. Any CPI measure of aggregate prices can be thought of as a weighted average over a set of K individual item indexes. Thus we can write:

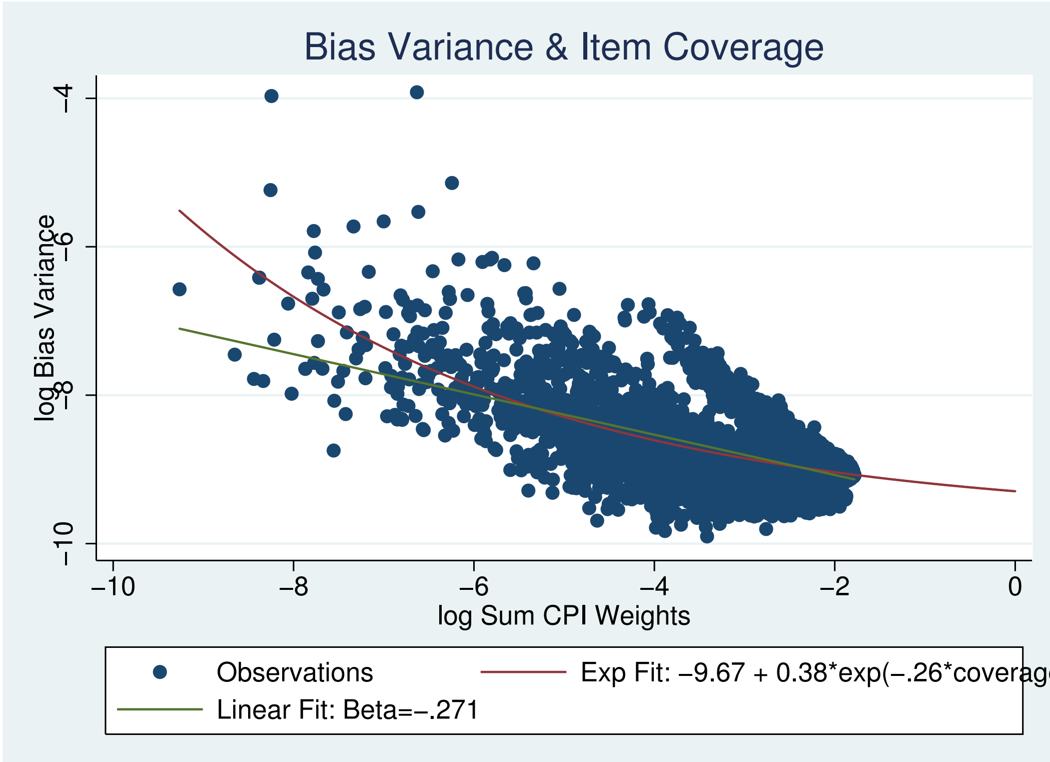
$$\pi_t^{CPI_K} = \sum_{i \in K} w_{i,t}^{CPI_K} \pi_{i,t}^{CPI}, \tag{12}$$

where $\pi_t^{CPI_K}$ is the CPI computed over set of K items, $w_{i,t}^{K}$ is the weight of item i in a CPI computed over set of K items, and $\pi_{i,t}^{CPI}$ is the CPI item index. Similarly, one can define a Törnqvist index over the same set of items:

$$\pi_t^{T_K} = \prod_{i \in K} \left(\pi_{i,t}^{T}\right)^{w_{i,t}^{T_K}} - 1, \tag{13}$$

where $\pi_t^{T_K}$ is a Törnqvist index computed over the K items, $w_{i,t}^{T_K}$ is the Törnqvist weight of item i in a Törnqvist index computed over the K items, and $\pi_{i,t}^{T}$ is the Törnqvist item index. We now can define the bias in the K item index as $\phi_t^K \equiv \pi_t^{CPI_K}$-$\pi_t^{T_K}$, and its variance across all time periods as $\sigma_{\phi_K}^2$. Similarly, we can define share of of the K items in the *full* CPI as $W_K = \frac{0.168}{T} \sum_{t \in T, i \in K} w_{i,t}^{CPI_K}$, where T is the number of time periods, and 0.168 is the share

Figure 10



of the total CPI expenditure categories in our data. For each value of K from one to 178, we can draw fifty samples with replacement and record the values of $\sigma^2_{\phi_K}$ and W_K.

Figure 10 plots the result of this exercise. Not surprisingly, we see that the bias variance declines with sample size, but what is striking is that the impact of increasing the sample size has a rapidly diminishing effect on reducing the variance. This is exactly what one might have expected given that some of the variance of the error is due to formula errors and not due to simply having a larger sample. Indeed, if we fit a Weibull function to this distribution, we estimate that using the full sample of all CPI products would only reduce the variance by 30 percent relative to our sample, and this result, of course, assumes that the sampling and formula errors in the non-grocery components of the CPI are not higher than those within the grocery sector. However, our simulation suggests that the mere fact that the CPI is computed over more items is unlikely to dramatically alter the fact that formula biases generate significant inflation measurement errors that cannot be averaged away.

6.2 Can time averaging reduce non-linearities?

Thus far, we have been arguing that the key problem in inflation measurement is the usage of formulas that do not correctly capture substitution of goods by consumers. Comprehensive overhaul of a country's price indexes is not easy, however, so one question that remains is whether there are any quick fixes that could mitigate this problem. One such approach is to use time averaging and compute inflation over longer periods: perhaps a 24-, 36-, or 48-month inflation rate will be more accurate than a 12-month one. Taking longer differences will not correct formula biases, but it can reduce sampling noise under certain circumstances.

For example, if inflation rates are correlated over time, *i.e.,* $Cov\left(\pi_t^T, \pi_{t-1}^T\right) > 0$, but the measurement errors are not, *i.e.,* $Cov\left(\phi_t^T, \phi_{t-1}^T\right) = 0$, we show in Appendix C that time averaging can increase the signal-to-noise ratio and reduce the bias. The intuition behind this result comes from the fact that if true inflation rates are positively correlated but the noise is not, then the noise will tend to cancel as we average over longer time periods, but the true inflation signal will amplify. For example, if the measured inflation rate for each month t in period 1 is $\pi_1^T + \phi_t$ and the measured inflation rate in in month s in period 2 is $\pi_2^T + \phi_s$, one could obtain a much better estimate of the difference in inflation rates between the two periods by comparing the difference in the average inflation rate in each period than by just computing the difference in inflation rates between the last month in period 1 and the first month in period 2.

We can see from Figure 2, that there is a strong reason to be believe that $Cov\left(\pi_t^T, \pi_{t-1}^T\right) > 0$ since inflationary and deflationary periods seem to be clustered together. However, when we look at Figure 5, we can also see that $Cov\left(\phi_t^T, \phi_{t-1}^T\right) > 0$, which suggests that the errors are also persistent. We can demonstrate this formally by regressing π_t^T on π_{t-1}^T and ϕ_t on ϕ_{t-1} to see if there is positive correlation. When we do this using Newey-West standard errors (with 11-lags), we find that coefficient from the first regression is 0.98 with a standard error of 0.03, and the coefficient from the second regression is 0.94 with a standard error of 0.03. In other words, while it is true that *month-to-month* measures of 12-month inflation are persistent, it is also true that the measurement errors in 12-month inflation estimates are almost equally persistent. The persistence of these errors is likely to diminish the advantages of time averaging because time averaging does not cause the errors to cancel (at least in monthly data).

One alternative would be to look at the persistence of inflation in one 12-month period ending in, say December, on the past year's 12-month inflation rate in December. When we do this we find that there is no significant correlation ($\rho = 0.2$) between the December 12-month inflation rate and its lag. Again, we find that time averaging is unlikely to help because there is little persistence in inflation rates across years.

Table 7
Impact of Measuring Inflation Over Longer Horizons

	No. of months used to compute inflation			
	12	24	36	48
Grocery CPI	0.553***	0.571**	0.621**	0.633***
	(0.151)	(0.221)	(0.241)	(0.198)
Grocery CPI^2	0.119**	0.164**	0.153*	0.110
	(0.0494)	(0.0753)	(0.0906)	(0.101)
Constant	-0.908***	-0.876***	-0.823***	-0.785***
	(0.224)	(0.263)	(0.225)	(0.190)
Implied Price Stability Target	1.285	1.152	1.052	1.049
Adjusted R^2	0.717	0.727	0.732	0.773

Newey-West standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Despite the fact that the data suggests that the theoretical conditions necessary to allow time averaging to improve the link between CPI inflation and true inflation are unlikely to hold, we can directly see the impact of time averaging in Table 7 where we explore how various forms of time averaging affect our estimates of β. Since most of the non-linear relationship between CPI inflation and actual inflation can be captured by the quadratic functional form (because there are not many periods of severe deflation in Japan), we will focus our results on this specification. As one can see, computing inflation over 24, 36, or 48 month periods does not tend to change the coefficients much, although the standard errors rise, presumably because longer differences require us to throw out more data.

7 Conclusion

This paper shows that the relationship between the economic concepts of inflation and the inflation indexes reported in official statistics is nonlinear. In particular, changes in the CPI and PCE deflator overstate changes in true inflation when inflation is low, and are only accurate measures when inflation is high. This result stems from the fact that much of the movement in the CPI (or our replicated PCE deflator) in low inflation regimes arises from formula errors in the computation of price indexes. Since the variance of true inflation is high in high-inflation regimes, the signal-to-noise ratio of the CPI rises with inflation making it more accurate in inflationary periods.

Moreover, our estimates suggest that the the biases of official inflation indexes are not constant. Inflationary biases peak when official inflation rates are at two percent per year, but approach zero at higher inflation rates. Thus, there is no single bias number for the CPI. Moreover, we are also able to show that even a chained official index, like the PCE deflator, has an upward bias relative to a Törnqvist index. This bias is largely due to the usage of unweighted geometric averages of price quotes at the lower level.

Our results have a number of implications for policy. First, given the biases inherent in the Japanese CPI, an inflation target of 2 percent approximately corresponds to price stability, *i.e.,* an inflation rate of zero. Similarly, our results suggest that price stability in the U.S. would correspond to a PCE deflator inflation rate of 0.5 percent. These results also suggest that the switch to geometric averaging at the lower level of the U.S. CPI and chaining at the upper level have not eliminated the upward bias in the PCE deflator or chained-CPI. This implies that indexing social security and other benefits to a chained index like the PCE or the chained-CPI may provide real income gains to recipients since the rate of increase in benefits will exceed that of true inflation.

Second, our results also imply that central banks would do well to pay less attention to official inflation measures in low inflation regimes and focus more on other variables that are better measured. Similarly, economists using official price indexes to measure inflation should be cognizant of the fact that these measures do not always move one to one with actual inflation.

References

Naohito Abe and Akiyuki Tonogi. Micro and macro price dynamics in daily data. *Journal of Monetary Economics*, 57(6):716–728, 2010.

Kosuke Aoki. On the optimal monetary policy response to noisy indicators. *Journal of Monetary Economics*, 50(3):501–523, 2003.

Laurence Ball. Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, 29(3):371–388, 1992.

Laurence Ball and Stephen G. Cecchetti. Inflation and uncertainty at short and long horizons. *Brookings Papers on Economic Activity*, 1990(1):215–254, 1990.

Laurence Ball, N. Gregory Mankiw, and David Romer. The new Keynesian economics and the output-inflation trade-off. *Brookings Papers on Economic Activity*, 1988(1):1–82, 1988.

Christian Broda and David E. Weinstein. Product creation and destruction: Evidence and price implications. *American Economic Review*, 100:691–723, 2010.

Stephen G. Cecchetti. Measuring short-run inflation for central bankers. *Review of the Federal Reserve Bank of St. Louis*, 79(3):143–156, 1997.

Alex Cukierman and Allan H. Meltzer. A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica*, 54(5):1099–1128, 1986.

W Erwin Diewert. Superlative index numbers and consistency in aggregation. *Econometrica*, pages 883–900, 1978.

W Erwin Diewert. The early history of price index research. In W Erwin Diewert and Alice O Nakamura, editors, *Essays in Index Number Theory*, volume 1, pages 33–65. North Hollands, Amsterdam, 1993. URL `http://EconPapers.repec.org/RePEc:nbr:nberwo:2713`.

W. Erwin Diewert and Alice O. Nakamura. *Essays on Index Number Theory: Volume 1*. North-Holland, 1993.

Stanley Fischer. Relative shocks, relative price variability, and inflation. *Brookings Papers on Economic Activity*, 1981(2):381–441, 1981.

Milton Friedman. Nobel lecture: Inflation and unemployment. *The Journal of Political Economy*, pages 451–472, 1977.

Etienne Gagnon. Price setting during low and high inflation: Evidence from mexico. *The Quarterly Journal of Economics*, 124(3):1221–1263, 2009.

Jerry Hausman. Sources of bias and solutions to bias in the consumer price index. *the Journal of Economic Perspectives*, 17(1):23–44, 2003.

Satoshi Imai, Chihiro Shimizu, and Tsutomu Watanabe. How fast are prices in japan falling? Discussion papers 12075, Research Institute of Economy, Trade and Industry (RIETI), 2012.

Michael T. Kiley. Endogenous price stickiness and business cycle persistence. *Journal of Money, Credit and Banking*, pages 28–53, 2000.

David E. Lebow and Jeremy B. Rudd. Measurement error in the consumer price index: Where do we stand? *Journal of Economic Literature*, 41(1):159–201, 2003.

Robert E Lucas. Expectations and the neutrality of money. *Journal of Economic Theory*, 4 (2):103–124, 1972.

Robert E. Lucas. Some international evidence on output-inflation tradeoffs. *American Economic Review*, 63(3):326–334, 1973.

Arthur M. Okun. The mirage of steady inflation. *Brookings Papers on Economic Activity*, 1971(2):485–498, 1971.

Richard W. Parks. Inflation and relative price variability. *The Journal of Political Economy*, pages 79–95, 1978.

Marshall Reinsdorf and Jack E. Triplett. A review of reviews: Ninety years of professional thinking about the consumer price index. In *Price Index Concepts and Measurement*, pages 17–83. University of Chicago Press, 2009.

Matthew D. Shapiro and David W. Wilcox. Mismeasurement in the consumer price index: An evaluation. In *NBER Macroeconomics Annual 1996, Volume 11*, pages 93–154. MIT Press, 1996.

David J. Stockton. Relative price dispersion, aggregate price movement, and the natural rate of unemployment. *Economic Inquiry*, 26(1):1–22, 1988.

Lars E.O. Svensson and Michael Woodford. Indicator variables for optimal policy. *Journal of Monetary Economics*, 50(3):691–720, 2003.

John B. Taylor. On the relation between the variability of inflation and the average inflation rate. In *Carnegie-Rochester Conference Series on Public Policy*, volume 15, pages 57–86. Elsevier, 1981.

Daniel R. Vining and Thomas C. Elwertowski. The relationship between relative prices and the general price level. *American Economic Review*, 66(4):699–708, 1976.

Appendix A Derivations

Variance of the CPI: The CPI is related to the Törnqvist index by the following equation:

$$\pi_t^{CPI} := \sum_i (w_{it} + \epsilon_{it})(\pi_{it} + \delta_{it}) = \pi_t^T + \sum_i w_{it}\delta_{it} + \sum_i \epsilon_{it}\pi_{it} + \sum_i \epsilon_{it}\delta_{it}. \tag{14}$$

By expanding (14), we obtain

$$\begin{aligned} V\left(\pi_t^{CPI}\right) &= V\left(\pi_t^T\right) + V\left(\sum_i w_{it}\delta_{it}\right) + V\left(\sum_i \epsilon_{it}\pi_{it}\right) + V\left(\sum_i \epsilon_{it}\delta_{it}\right) \\ &+ 2Cov\left(\pi_t^T, \sum_i w_{it}\delta_{it}\right) + 2Cov\left(\pi_t^T, \sum_i \epsilon_{it}\pi_{it}\right) + 2Cov\left(\pi_t^T, \sum_i \epsilon_{it}\delta_{it}\right) \\ &+ 2Cov\left(\sum_i w_{it}\delta_{it}, \sum_i \epsilon_{it}\pi_{it}\right) + 2Cov\left(\sum_i w_{it}\delta_{it}, \sum_i \epsilon_{it}\delta_{it}\right) + 2Cov\left(\sum_i \epsilon_{it}\pi_{it}, \sum_i \epsilon_{it}\delta_{it}\right). \end{aligned}$$

Notice that all the covariance terms are zero because the error terms are independent. For example, consider the first covariance term. We know that the expected value of the weighted average of the error terms is zero because

$$E\left[\sum_i w_{it}\delta_{it}\right] = \sum_i E\left[w_{it}\delta_{it}\right] = \sum_i E\left[w_{it}\right] E\left[\delta_{it}\right] = 0$$

which gives

$$\begin{aligned} nCov\left(\pi_t^T, \sum_i w_{it}\delta_{it}\right) &= E\left[\left(\sum_i w_{it}\pi_{it}\right)\left(\sum_i w_{it}\delta_{it}\right)\right] = E\left[\sum_{i,j} w_{it}w_{jt}\pi_{it}\delta_{jt}\right] \\ &= \sum_{i,j} E\left[w_{it}w_{jt}\pi_{it}\right] E\left[\delta_{jt}\right] = 0. \end{aligned}$$

A similar argument holds for all other covariance terms. Therefore,

$$V\left(\pi_t^{CPI}\right) = V\left(\pi_t^T\right) + V\left(\sum_i w_{it}\delta_{it}\right) + V\left(\sum_i \epsilon_{it}\pi_{it}\right) + V\left(\sum_i \epsilon_{it}\delta_{it}\right). \tag{15}$$

We can write the second term of (15) in terms of the underlying variances of the errors through a bit of algebra. First recall $E\left[\sum_i w_{it}\delta_{it}\right] = \sum_i E\left[w_{it}\delta_{it}\right] = \sum_i E\left[w_{it}\right] E\left[\delta_{it}\right] = 0.$

Now we can write,

$$
\begin{aligned}
V\left(\sum_i w_{it}\delta_{it}\right) &= E\left[\left\{\frac{1}{n}\sum_{i,j}\left(s_{it-1}+\frac{\gamma}{2}\left(\nu_{it}-\nu_{jt}\right)\right)\delta_{it}\right\}^2\right] \\
&= \frac{1}{n^2}\sum_{i,j,k}E\left[\left(s_{it-1}+\frac{\gamma}{2}\left(\nu_{it}-\nu_{jt}\right)\right)\left(s_{it-1}+\frac{\gamma}{2}\left(\nu_{it}-\nu_{kt}\right)\right)\delta_{it}^2\right] \quad (\because \delta_{it}\perp\delta_{jt} \text{ if } i\neq j) \\
&= \frac{\sigma_{\delta_t}^2}{n^2}\sum_{i,j,k}E\left[\left(s_{it-1}^2+\frac{\gamma^2}{4}\left(\nu_{it}-\nu_{jt}\right)\left(\nu_{it}-\nu_{kt}\right)\right)\right] \quad (\because E\left[\nu_{it}-\nu_{kt}\right]=0\forall i,k) \\
&= \frac{\sigma_{\delta_t}^2}{n^2}\left(\sum_{i,j,k}s_{it-1}^2+\frac{\gamma^2}{4}\left(\sum_{i,j,k}\underbrace{E\left[\nu_{it}^2\right]}_{=\sigma_{\nu_t}^2}-\sum_{i,j,k}\underbrace{E\left[\nu_{it}\nu_{kt}\right]}_{=\sigma_{\nu_t}^2 1_{\{i=k\}}}-\sum_{i,j,k}\underbrace{E\left[\nu_{jt}\nu_{it}\right]}_{=\sigma_{\nu_t}^2 1_{\{j=i\}}}+\sum_{i,j,k}\underbrace{E\left[\nu_{jt}\nu_{kt}\right]}_{=\sigma_{\nu_t}^2 1_{\{j=k\}}}\right)\right) \\
&= \frac{\sigma_{\delta_t}^2}{n^2}\left(\sum_i s_{it-1}^2\sum_{j,k}1+\frac{\gamma^2}{4}\sigma_{\nu_t}^2\left(\sum_{i,j,k}1-\sum_{i,j,k}1_{\{i=k\}}-\sum_{i,j,k}1_{\{j=i\}}+\sum_{i,j,k}1_{\{j=k\}}\right)\right) \\
&= \frac{\sigma_{\delta_t}^2}{n^2}\left(n^2\sum_i s_{it-1}^2+\frac{\gamma^2}{4}\sigma_{\nu_t}^2\left(n^3-n^2-n^2+n^2\right)\right) \\
&= \sigma_{\delta_t}^2\left(\sum_i s_{it-1}^2+\frac{\gamma^2}{4}\left(n-1\right)\sigma_{\nu_t}^2\right).
\end{aligned}
$$

We can rewrite the third term of (15) in terms of the underlying variances by first remembering that $E\left[\sum_i \epsilon_{it}\pi_{it}\right]=\sum_i E\left[\epsilon_{it}\right]E\left[\pi_{it}\right]=0$. The independence of ϵ_{it} guarantees that the cross terms disappear, and we get

$$
\begin{aligned}
V\left(\sum_i \epsilon_{it}\pi_{it}\right) &= E\left[\left(\sum_i \epsilon_{it}\pi_{it}\right)^2\right]=\sum_i E\left[\left(\epsilon_{it}\pi_{it}\right)^2\right]=\sum_i E\left[\epsilon_{it}^2\right]E\left[\pi_{it}^2\right] \\
&= n\sigma_{\epsilon_t}^2\left(\sigma_{\mu_t}^2+\sigma_{\nu_t}^2\right).
\end{aligned}
$$

Finally, since the error terms are iid when $i\neq j$ and $E\left[\epsilon_{it}\right]=E\left[\delta_{it}\right]=0$, the fourth term of (15) can be rewritten as

$$
V\left(\sum_i \epsilon_{it}\delta_{it}\right)=n\sigma_{\epsilon_t}^2\sigma_{\delta_t}^2.
$$

In summary,

$$
\begin{aligned}
V\left(\pi_t^{CPI}\right) &= V\left(\pi_t^{T}\right) + \sigma_{\delta_t}^2\left(\sum_i s_{it-1}^2 + \frac{\gamma^2}{4}(n-1)\,\sigma_{\nu_t}^2\right) + n\sigma_{\epsilon_t}^2\left(\sigma_{\mu_t}^2 + \sigma_{\nu_t}^2\right) + n\sigma_{\epsilon_t}^2\sigma_{\delta_t}^2 \\
&= V\left(\pi_t^{T}\right) + \sigma_{\delta_t}^2\left(\sum_i s_{it-1}^2 + \frac{\gamma^2}{4}(n-1)\,\sigma_{\nu_t}^2\right) + n\sigma_{\epsilon_t}^2\left(\sigma_{\mu_t}^2 + \sigma_{\nu_t}^2 + \sigma_{\delta_t}^2\right)
\end{aligned}
$$

Covariance of the CPI and true inflation: Note $Cov\left(\pi_t^T, \sum_i w_{it}\delta_{it}\right) = Cov\left(\pi_t^T, \sum_i \epsilon_{it}\pi_{it}\right) = Cov\left(\pi_t^T, \sum_i \epsilon_{it}\delta_{it}\right) = 0$. Hence,

$$
\begin{aligned}
Cov\left(\pi_t^{CPI}, \pi_t^T\right) &= V\left(\pi_t^T\right) + Cov\left(\pi_t^T, \sum_i w_{it}\delta_{it}\right) + Cov\left(\pi_t^T, \sum_i \epsilon_{it}\pi_{it}\right) + Cov\left(\pi_t^T, \sum_i \epsilon_{it}\delta_{it}\right) \\
&= V\left(\pi_t^T\right).
\end{aligned}
$$

Appendix A.1 Decomposing CPI Measurement Error

We can use the upper-level components of the item Törnqvist to study the measurement error in the Japanese CPI through the lens of the theoretical framework presented in Section 3 above. In this section, we derived the relationship between the aggregate measurement error in the Japanese CPI and the measurement errors in the underlying item-level components of this index (Equation 9). We can now define these errors in the weight and item inflation index components empirically as

$$
\varepsilon_{i,t}^{IT} = w_{i,t}^{CPI} - w_{i,t}^{T} \tag{16}
$$

$$
\delta_{i,t}^{IT} = \pi_{i,t}^{CPI} - \pi_{i,t}^{T} \tag{17}
$$

The relationship between these errors and the aggregate measurement error defined in Equation 9 also depended on the aggregate and idiosyncratic components of the true item level inflation. We extract these components by running the following unweighted regression: $\pi_{i,t}^T = \mu_t + \nu_{i,t}$.

In Section 3, we additionally formally characterized the measurement error in the Japanese

CPI. We found that the conditional expectation of true inflation based on observed inflation depends crucially on the variances of the inflation components (μ_t, ν_{it}) and errors (ϵ_t, δ_{it}) listed above. To characterize the inference problem empirically, we estimate these variances. We define the variance of the aggregate component of inflation, μ_t as

$$\hat{\sigma}^2_{\mu_t} = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{\mu}_t - \bar{\hat{\mu}}\right)^2, \text{ where } \bar{\hat{\mu}} = \frac{1}{N}\sum_{t=1}^{N}\hat{\mu}_t \tag{18}$$

and where N is the number of observations (months) in the sample. We then calculate the variances of ε, δ, ν within each item as

$$\hat{\sigma}^2_{x_{i,t}} = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{x}_{i,t} - \bar{\hat{x}}_i\right)^2, \text{ where } \bar{\hat{x}}_i = \frac{1}{N}\sum_{t=1}^{N}\hat{x}_{i,t} \tag{19}$$

where $x = \varepsilon, \delta, \nu$ specifies the variable under consideration. We aggregate these variances across items using Törnqvist item weights:

$$\hat{\sigma}^2_{x_t} = \sum_{i=1}^{I} w^T_{i,t}\hat{\sigma}^2_{x_{i,t}}$$

We use the values of each of these estimated variances to solve for the implied value of γ_t by equating

$$\frac{Cov\left(\pi^T_t, \pi^{CPI}_t\right)}{Var\left(\pi^{CPI}_t\right)} = \frac{Var\left(\pi^T_t\right)}{Var\left(\pi^T_t\right) + \sigma^2_{\delta_t}\sum_{i=1}^{n} s^2_{it-1} + n\left[\sigma^2_{\epsilon_t}\sigma^2_{\mu_t} + \sigma^2_{\epsilon_t}\sigma^2_{\nu_t} + \sigma^2_{\epsilon_t}\sigma^2_{\delta_t} + \frac{\gamma^2_t}{4}\left(\sigma^2_{\delta_t}\sigma^2_{\nu_t}\right)\right]} \tag{20}$$

using variables' period t values then solving for γ_t.

Appendix B Index Calculation

Our price and quantity data are at a daily frequency, so we sum over quantity and expenditure to find their monthly values. We then back out the monthly price by dividing. For all days

D_t in the month t, in each store s and JAN code j, we define

$$p_{j,s,t} = \frac{\sum_{d \in D_t} p_{j,s,d} q_{j,s,d}}{\sum_{d \in D_t} q_{j,s,d}} \tag{21}$$

We aggregate the daily item purchase and price data to a monthly frequency denoting the sales-weighted average price charged by store s for JAN code j in month t as $p_{j,s,t}$ and the corresponding sales quantity $q_{j,s,t}$. In order to do the decompositions indicated in equation (4), we need to work with Törnqvist index that is composed of Törnqvist item indexes. Each "item" level index is a Törnqvist index of the twelve-month within-store JAN code price changes aggregated across all JAN codes that belong to an "official" item category. The item-level Törnqvist index for item i in month t, $\pi^T_{i,t}$, is defined as follows:

$$1 + \pi^T_{i,t} = \prod_{j \in J_{i,s,[t-12,t]}} \prod_{s \in S_{[t-12,t]}} \left(\frac{p_{j,s,t}}{p_{j,s,t-12}} \right)^{w^T_{j,s,t}}$$

where $J_{i,s,[t-12,t]}$ is the set of item i JAN codes sold in store s in months $t-12$ and t; $S_{[t-12,t]}$ is the set of stores in the sample for months $t-12$ and t; and $w^T_{j,s,t}$ is item Törnqvist weight for JAN code j in store s, and month t given by

$$w^T_{j,s,t} = \frac{s_{j,i,s,t} + s_{j,i,s,t-12}}{2} \quad \text{for} \quad s_{j,i,s,t} = \frac{p_{j,s,t} q_{j,s,t}}{\sum_{j \in J_{i,s,[t-12,t]}} \sum_{s \in S_{[t-12,t]}} p_{j,s,t} q_{j,s,t}}$$

The item Törnqvist index aggregates these indexes across items $i = 1, ..., I$, using a weighted geometric average:

$$1 + \pi^T_t = \prod_{i=1}^{I} \left(1 + \pi^T_{i,t} \right)^{w^T_{i,t}} \tag{22}$$

with Törnqvist item weights, $w_{i,t}^T$ given by

$$w_{i,t}^T = \frac{s_{i,t} + s_{i,t-12}}{2} \quad \text{for} \quad s_{i,t} = \frac{\sum_{j \in J_{i,s,[t-12,t]}} \sum_{s \in S_{[t-12,t]}} p_{j,s,t} q_{j,s,t}}{\sum_{i=1}^{I} \sum_{j \in J_{i,s,[t-12,t]}} \sum_{s \in S_{[t-12,t]}} p_{j,s,t} q_{j,s,t}}.$$

[22]

Appendix C Does Averaging Reduce Noise When the Errors are Positively Correlated?

Suppose π^T is an AR(1) process, which is given by

$$\pi_t^T = a\pi_{t-1}^T + u_t \tag{23}$$

where a is a positive parameter $(0 < a < 1)$ and u_t is an iid disturbance term with a zero mean and a variance of σ^2. We define a new variable $\Pi_\tau^T(m)$ as

$$\Pi_\tau^T(m) \equiv \frac{1}{m} \sum_{t=m(\tau-1)+1}^{m\tau} \pi_t^T \tag{24}$$

for $\tau = 1, 2, \ldots$. The parameter m represents the order of aggregation. For example, when $m = 2$, we are averaging over two time periods and have

$$\Pi_1^T(2) \equiv \frac{1}{2} \sum_{t=1}^{2} \pi_t^T; \quad \Pi_2^T(2) \equiv \frac{1}{2} \sum_{t=3}^{4} \pi_t^T; \quad \Pi_3^T(2) \equiv \frac{1}{2} \sum_{t=5}^{6} \pi_t^T \quad ; \quad \ldots \tag{25}$$

The variance of Π^T is given by

$$Var\left(\Pi_\tau^T(m)\right) = \frac{m + \sum_{k=1}^{m-1} 2(m-k)a^k}{m^2} \frac{\sigma^2}{1-a^2} \tag{26}$$

[22]In the theory section, we make the approximation that $\ln\left(1 + \pi_t^T\right) = \pi_t^T$, but equation (22) does not make this approximation. This has virtually no impact on the results.

Similarly, the average measurement error is given by

$$\Phi_\tau(m) \equiv \frac{1}{m} \sum_{t=m(\tau-1)+1}^{m\tau} \phi_t \tag{27}$$

If we assume that ϕ_t is iid, the variance of $\Phi_\tau(m)$ is given by

$$Var(\Phi_\tau(m)) = \frac{\sigma_\phi^2}{m} \tag{28}$$

The signal-to-noise ratio can now be written as

$$\frac{Var\left(\Pi_\tau^T(m)\right)}{Var(\Phi_\tau(m))} = \frac{m + \sum_{k=1}^{m-1} 2(m-k)a^k}{m} \frac{\sigma^2}{\sigma_\phi^2(1-a^2)},$$

We can show that this is increasing in m by defining $A(m)$ as

$$A(m) \equiv \frac{m + \sum_{k=1}^{m-1} 2(m-k)a^k}{m} \tag{29}$$

It is useful to rewrite $A(m)$ as

$$A(m) = \frac{m + 2(m-1)a + 2(m-2)a^2 + \cdots + 2 \times 2 \times a^{m-2} + 2 \times 1 \times a^{m-1}}{m} \tag{30}$$

Then, we have

$$(m+1)A(m+1) = m + 1 + 2(m+1-1)a + 2(m+1-2)a^2 + \cdots + 2 \times 2 \times a^{m+1-2} + 2 \times 1 \times a^{m+1-1} \tag{31}$$

$$mA(m) = m + 2(m-1)a + 2(m-2)a^2 + \cdots + 2 \times 2 \times a^{m-2} + 2 \times 1 \times a^{m-1} \tag{32}$$

which implies

$$(m+1)A(m+1) - mA(m) = 1 + 2a + 2a^2 + \cdots + 2a^{m-1} + 2a^m. \tag{33}$$

Using equation (33), we have

$$(m+1)A(m+1) - (m+1)A(m) = 2\left(\frac{1}{m}a + \frac{2}{m}a^2 + \cdots + \frac{m-1}{m}a^{m-1} + \frac{m}{m}a^m\right). \qquad (34)$$

Since the right-hand side of this equation is positive, we have

$$A(m+1) > A(m), \qquad (35)$$

which means that the signal-to-noise ratio must be rising with averaging and hence β will also rise.