BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

Combining Hedonic with Multilateral Indexes for Turnover and Chain Drift in Transactions Data Consumer Price Indexes

Gregory Kurtzon,
U.S. Bureau of Labor Statistics

Working Paper 587
September 2025

Combining Hedonic with Multilateral Indexes for Turnover and Chain Drift in Transactions Data Consumer Price Indexes

Gregory Kurtzon

September 2025

Abstract

This paper combines hedonics and rolling window multilateral indexes in transactions data to fix the problems of product turnover and chain drift. Transactions data hold the potential to eliminate substitution bias in elementary price indexes with quantities concurrent with prices, but using these quantities can cause chain drift (non-circularity). The typical solution in national consumer price indexes (CPIs) is to use multilateral indexes, but these don't fully eliminate chain drift. The large amount of product entry and exit is usually delt with by making unit value prices, but this causes biases from quality changes and Jensen's inequality. Previous research has studied different multilateral formulas, window lengths, and rolling window splicing methods, and noted that sales and product turnover are highly connected to chain drift. Therefore, hedonic imputation should help all these issues at the same time. I create Pakes (2003) style hedonic indexes using IRI grocery and drug store data from 2001-2011 with dummy variables for all variable values plus interactions and bootstrap the indexes to check for overfitting. I make Caves, Christensen, and Diewert (1982) (CCDI) indexes over rolling window lengths that span the entire range of data and make Multiperiod Identity Tests and Laspeyres-Paasche spreads to evaluate them. Chain drift is typically improved for short rolling windows, but doesn't disappear, and can actually get better or worse for longer windows until the window length converges to all twelve years. The hedonic indexes have an acceptable level of variance and their use to some extent has the potential to reduce chain drift themselves by smoothing the effects of sales.

JEL Codes: C43, E31, D44, D12

1 Introduction

Transactions data is now being widely demanded and used in official national CPIs in many major nations. The promise of these new data sources is the ability to avoid possible sampling biases from using the whole universe of transactions, and also to avoid substitution bias by using quantities concurrent with prices to make monthly chained indexes. However, there are two major problems with using this data. One is that there's a large number of new and entering goods, and with the whole universe already in the sample, there is no room to use the typical methodology of substituting a good for an exiting one. Another problem is that using the raw concurrent quantities gives the index chain drift, here defined as non-circularity, meaning that even if all prices return to their original levels, the chained index level will not return to 1 as it should. This paper studies the use of hedonics to deal with both good turnover and non-circularity, and its combination with multilateral indexes that are typically used to treat non-circularity.

The method typically used to incorporate new and exiting goods by those countries that use a large amount of transactions data is to make a unit value index over a thin band of goods' characteristics. However, unit value indexes will treat all quality improvements in new goods as price increases, and will also suffer from Jensen's inequality, $E\left(\frac{p_{it}}{p_{i,t-1}}\right) \leq \frac{E(p_{it})}{E(p_{i,t-1})}$, so they would suffer upward bias. Unit value indexes are also not justified by economics and often have erratic behavior. Making a pure hedonic index, as per Pakes (2003) however, would solve these problems and be economically justified.

Chain drift is typically caused by price bouncing. Price index formulas are approximations for an arbitrary expenditure function, derived under the conditions that quantities are consumed in the period they are purchased. In fact, they are all discrete time approximations the same continuous Divisia index which is a the Cost-of-Living index for any consumer preferences.[1] All these approximations will work well close the point of expansion which is when the prices between any two periods in a chain link are equal, and if consumers do not stockpile goods. If the prices converge as the interval grows shorter, these approximations are very similar, as shown in Kurtzon (2022). However, if prices bounce up and down between periods, and if consumers stockpile goods, chain drift can occur. For example, during a sale the price falls a large amount for one period, and the quantity is very high as consumers stock up. After the sale, the price returns but the quantity may be lower than before since consumers are consuming stored goods. A Törnqvist index will have a lower mean share for the upward price movement than the lower one, and the index will drift downward despite the prices returning, and so the index will be non-circular. Other patterns in quantities can also occur, as described in Hill (2006a), which have different biases in indexes.

However, how to measure chain drift isn't clear. If we had an index without chain drift to measure a proposed index against, a true measure as a gold standard, we wouldn't need another index and there would have been no problem in the first place. But of course there is no measure that we can be sure isn't biased. Sometimes, such as noted in Ivanic, Diewert, and Fox (2011) (IDF), one can look at an index and see that a 20% deflation rate for some goods in a

[1] See Balk (2005), Diewert (1976), Divisia (1926), Reinsdorf (1998), CPI Manual ILO 2004.

weekly aggregated index, is 'unreasonable', using priors on how prices behave. This 'reasonability' test is very vague and informal, but can be a kind of last resort for extreme indexes.

Or, using the definition of chain drift, the Multiperiod Identity Test (MPID test), invented by Walsh (1901) (referred to as the MPID test in Diewert (1993)), measures the degree of non-circularity by the difference between an index with all prices returning to their first value, and 1. To make an index with all prices returning to the first value, an additional period is added to a chained index, with the base of the last period and the current period being the first.

But a problem is that the last chain link in the MPID test will span a much longer period of time than the typical link, and the prices from period 1 to T will be more dissimilar, meaning the index will be a worse approximation of the expenditure function and of true consumer preferences. Hill (2006b) for example, shows that if long time spans for index links are combined with short time spans, even superlative indexes, which are second order approximations to arbitrary preferences[2], will diverge from each other, showing that they become worse approximations. Another problem is that over longer periods of time there is more turnover in goods - there are fewer goods sold in both 1 and T than between other links t-1 to t, which also makes the last link less reliable. Appendix 7.1 discusses an alternate measure of chain drift, using the spread between GEKS Laspeyres and Paasche indexes.

While there have been many other papers on chain drift and multilateral indexes (including but of course not limited to Diewert and Fox (2022), Fox, Levell, and O'Connell (2023), Feenstra and Shapiro (2003), Melser and Webster (2021)), and on hedonic price indexes, there has been little on the combination of them. The time dummy characteristics hedonic index (or time dummy hedonics index) is a multilateral method that is also a hedonic index that uses good's characteristics, but it has the limitation that the characteristics coefficients are held constant over the whole multilateral window. Since the main bias of excluding new and entering goods, as described in Pakes (2003) and Erickson and Pakes (2011), is that characteristics coefficients change, this can lose much of the bias reduction of a hedonic index. Since a hedonic index can be thought of as simply redefining goods as characteristics, this is similar to holding prices constant. Diewert and Shimizu (2024) study this and limit the bias from holding characteristics prices constant by using only two adjacent periods at a time, which is a very short rolling window subject to more chain drift, and compare it to an expanding window which uses all available data but holds characteristics prices constant longer. DeHaan & Krisinich (2014) also study monthly full imputation hedonic indexes in a rolling window Caves, Christensen, and Diewert (1982) (CCDI) (or Törnqvist GEKS) multilateral index.

However, the main contributions of this paper are (1) the bootstrapping of full imputation hedonic indexes with dummy variables, to show variance and strength of the specification; (2) showing the combination of full imputation

[2]See Diewert (1976).

indexes and a multilateral of all window lengths for a long period of time, and how it converges to circularity with a MPID test; (3) the chain drift reducing effects of hedonic indexes. Some amount of hedonic imputation is shown to have the potential on its own to reduce chain drift. However, a multilateral index such as the CCDI may improve drift or may make it worse depending on the window length.

The rest of the paper is as follows. Section 2 describes the hedonic regression and bootstrapping method, the CCDI multilateral index, and the Multiperiod Identity Test. Section 3 describes the data and results. Section 4 concludes. The Appendices describe alternative methods for measuring and reducing chain drift.

2 Hedonics for Transactions Data

Hedonic indexes were constructed using the methodology of Pakes (2003), where a separate regression is run for every month. The strength of this method is that there are no restrictions holding the coefficients constant over time, and only information from each period alone is used to make predictions. If characteristics prices change significantly over a month, this method is more accurate. To use the maximum amount of information available, one dummy variable was used for every possible value of each of the goods' characteristics. For example, for cold cereal, there is a dummy variable for every store, a dummy for every value of "Type of Grain", including a dummy for "Assorted", "Multi Grain", "Corn", etc., and similarly for the other variable's values.

Chained Törnqvist indexes were made using the predicted prices and monthly quantities. Two types of imputed prices were made: non-missing only and full imputation. The first replaces all observed prices with a predicted price from the hedonic regression, replacing every observed price for item i in period t, p_{it}, with the predicted price $\hat{p}_{it}$. This type of imputation helps study the effects of hedonics without it's effect on entering and exiting goods. The second method imputes all missing prices, so that the predicted prices from each regression were used in place of both actual and missing prices for every good that was ever in the sample. For periods that the store for a given missing good is not in the sample, it simply sets the store dummy coefficient for period t to zero, $\beta_{store,t} = 0$. When a good exits or enters, the share for the month for a good that was not present was therefore zero, and the Törnqvist averages it with the share of the good after it enters or before it exits. Thus, there is a price relative that includes all entering and exiting goods to avoid the selection biases of quality changes. When for two consecutive periods a good is not sold, the average share is zero, and so it is effectively excluded.

The basic model for prices p_{it} for period t and item, dummy variable D_{ci} which is one if item i has characteristic c and zero otherwise, characteristic coefficient β_{ct}, and error ε_{it} is

$$p_{it} = \left(\Pi_c \beta_{ct}^{D_{ic}}\right) \varepsilon_{it} \tag{1}$$

while the regression model run is

$$\ln p_{it} = \Sigma_c D_{ci} \beta_{ct} + \ln \varepsilon_{it} \tag{2}$$

. The regression is run on log prices so that the error term is multiplicative and does not bias a price relative. But because $\ln \varepsilon_{it}$ is approximately normally distributed, the expectation of price indexes made with the antilog is

$$\exp\left(\widehat{\ln p_{it}}\right) = \exp\left(\Sigma_c D_{ci} \hat{\beta}_{ct} + \widehat{\ln \varepsilon_{it}}\right) = \exp\left(\Sigma_c D_{ci} \hat{\beta}_{ct}\right) \exp\left(\widehat{\ln \varepsilon_{it}}\right) \tag{3}$$

. Because $\ln \varepsilon_{it}$ is lognormally distributed, $\exp\left(\widehat{\ln \varepsilon_{it}}\right)$ will not have an expectation of 1, but instead of $\exp\left(\frac{1}{2}\sigma^2\right)$, where σ^2 is the variance of the distribution of ε_{it}. Therefore a Poisson regression is used to estimate the model and generate predicted prices $\hat{p}_{it}$. The Poisson regression is a nonlinear regression originally designed for count data, but does not have biased predictions. The regression is weighted by quantity, so that an item that is sold multiple times will carry the same importance as multiple items that sold only one unit.

The hedonic Törnqvist index between periods t-1 and t, where the expenditure share of item i in period it is s_{it}, is

$$T_{t-1,t}^{H} = \Pi_i \left(\frac{\hat{p}_{it}}{\hat{p}_{i,t-1}}\right)^{\frac{s_{i,t-1}+s_{it}}{2}} \tag{4}$$

and the index level is

$$I_t^H = \Pi_{\tau=1}^{t} T_{\tau-1,\tau}^{H} \tag{5}$$

With dummy variables for every characteristic there are a very large number of variables[3], so overfitting is a possibility that needs to be tested for these regressions. Overfitting occurs when so many degrees of freedom are used by the variables that the coefficients and predictions have such a high variance that they are can't be relied on for a certain purpose. If variance is a problem, the typical solution is to use one of many methods (including "machine learning" methods) to reduce variance at the cost of additional bias. But as Pakes (2003) points out, the predictions for prices only matter for constructing indexes.

Bias becomes more important over time than variance for price indexes, since variance in the relatives averages out over time for the levels, while bias accumulates. If variance in the index levels is below an acceptable threshold for a time period when they are needed, reducing bias should take precedence over variance.

To measure this, fifty bootstraps of the index levels were done for every goods category. For each bootstrap, all items that were ever in the sample were

[3] Perfectly colinear variables are automatically removed for any month's regression.

randomly sampled with replacement, then a separate regression was run every month with the bootstraped sample of items, and a new hedonic index series was made. The mean and standard deviation of these index levels was calculated for each period, and a near 95% (95.4%) confidence interval was made by adding and subtracting two standard deviations from the mean for each month.

2.1 The CCDI index and MPID Test

The CCDI method is a multilateral method often used with transactions data. For a given window of periods T, it makes unnormalized index levels ρ_t for each period in the window by taking the average Törnqvist index for that period and every other period in the window, and then normalizing the levels by the first period in the data to make the index level π_t,

$$\rho_t = \left[\Pi_{\tau=1}^{T}\Pi_i \left(\frac{p_{it}}{p_{i\tau}}\right)^{\frac{1}{2}(s_{i\tau}+s_{it})}\right]^{\frac{1}{T}} \tag{6}$$

$$\pi_t = \frac{\rho_t}{\rho_1} \tag{7}$$

This is the same as the GEKS (Gini (1931); Elteto and Koves (1964); Szulc (1964)) index, except that it uses the Törnqvist formula instead of the Fisher formula for bilateral indexes.

This will be exactly circular if the entire data range is used. However, for a long time span, this means they will be using long direct indexes as inputs that are poor Divisia approximations and have fewer goods common between the base and current periods, either losing goods or requiring more price predictions of missing prices. The common solution used by statistical agencies is a rolling window index, as proposed by Ivanic, Diewert, and Fox (2011). Every period the base period moves forward, so that the CCDI indexes are made for a window of the same length. The index change in the new window for the latest month must then be "spliced" onto a index level of the old window for a continuous index series. Unnormalized indexes levels are made for each window, and then spliced together to form the final, normalized index series. Many splicing methods have been proposed and studied, and depending on the data, different methods may nor may not make a significant difference. Fox, Levell, and O'Connell (2024) show that in practice they may make little difference, while von Auer (2024), Chessa (2025), and others show they can be significant. The simplest method is the end splice, which simply takes the last index relative for the latest window and splices it onto the last normalized level. If the splicing months are very different, using the mean over all months, the mean splice, recommended by von Auer (2024), or a simpler approximation, the median splice, could be better. Appendix 3 shows evidence that the results change little in this context if an end splice or half splice is used. The CCDI index levels using the end splice and half splice methods are shown on the same graphs for comparison. The index levels are almost the same for the two methods, with the half splice showing slightly more variability across window lengths.

Therefore the following results present the half splice for its combination of simplicity and robustness, which takes the inflation rate from the median to the end of the most recent window and multiplies (or splices) it onto the last window's median normalized rolling window level. The normalized index level π_t for period t, splicing in period $\frac{l+1}{2}$, denoting the unnormalized index levels be denoted $\rho^w_{l,t}$ for window number w of length l in period t is

$$\pi_t = \pi_{t-\frac{l+1}{2}+1} \quad \left. \frac{\rho^w_{l,t}}{\rho^w_{l,t-\frac{l+1}{2}+1}} \right) \tag{8}$$

where the asterisk (*) denotes the unnormalized levels constructed from the new rolling window.

The multiperiod identity (MPID) test is used as the measure of chain drift. Let $P\left(p^s, p^r, q^s, q^r\right)$ denote the price index between periods s and r with prices p^s and p^r and quantities q^2 and q^r. For a chained index of T periods,

$$MPID \text{ test value} = P\left(p^1, p^2, q^1, q^2\right) P\left(p^2, p^3, q^2, q^3\right) \cdots P\left(p^{T-1}, p^T, q^{T-1}, q^T\right) P\left(p^T, p^1, q^T, q^1\right) \tag{9}$$

which should equal 1. If not, the index is said to have chain drift under the definition which is non-circularity. This was made for the chained Törnqvist indexes for the non-missing only imputation, full hedonic imputation, and raw prices indexes.

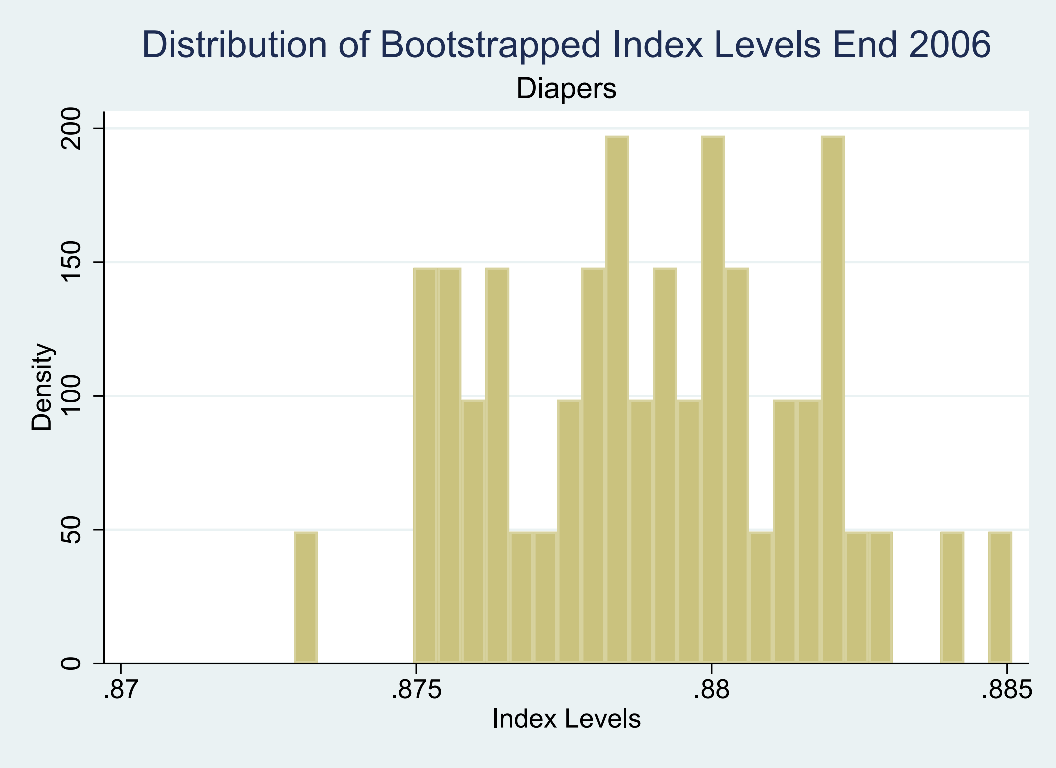
3 Data and Results

The data is the IRI Academic Dataset for 2001 through 2012. It contains 38 goods categories of grocery and drug store sales, from all large retail chains, in weekly unit value prices and total quantities for each UPC of goods. This paper focuses on only 14 of those goods, half of them food: cold cereal, carbonated beverages, coffee, deodorant, diapers, facial tissue, laundry detergent, mayonnaise, peanut butter, paper towels, razors, salty snacks, toothbrushes, and yogurt. In order to avoid using parts of the weekly unit value to construct monthly indexes, each four week period was aggregated to make 13 artificial 'months' for each year. IRI constructs enough characteristics variables, including a store identifier, to identify every good at each store it is sold at. For example, the cold cereal category has characteristics for store, volume equivalent, flavor/scent, sugar content, package, fiber info, fortification, and type of grain, and has about 1.5 million observations per 4-week 'month'.

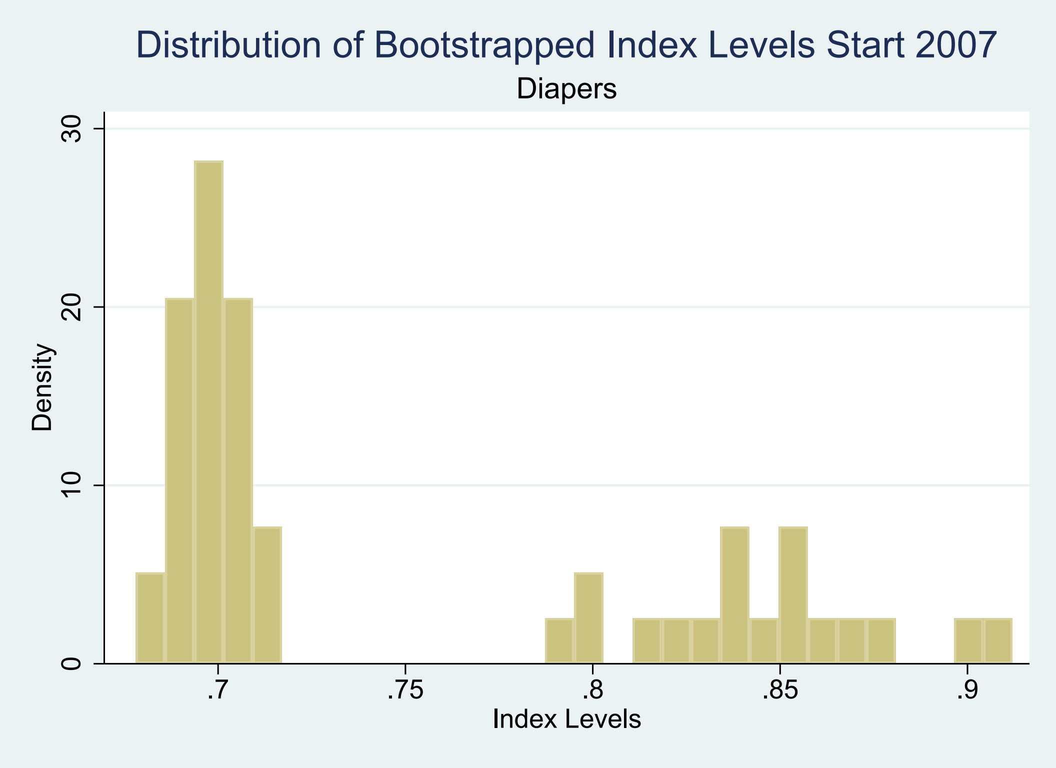
Appendix 1 shows the mean bootstraped index and confidence interval for the 14 goods categories[4]. About half have very tight intervals for all months, carbonated beverages, cold cereal, peanut butter, yogurt, deodorant, mayonnaise, and laundry detergent. For the other categories, the intervals are tight

[4] Weighted Poisson bootstrapped regressions did not converge for carbonated beverages, coffee, cold cereal, salted snacks, tooth brushes, and yogurt, so the results presented are for unweighed linear regressions.

for almost all months, except for one or two months, the first 4 week 'month' of the year, often the same across these categories, for 2007, 2008, and 2012, where the bootstraped indexes diverge. The variance makes a large jump, the distribution of indexes becomes bimodal, and the mean shifts. This divergent month is often the same across these categories, such as month 79 (first month of 2007) for diapers, facial tissue, coffee, and toothbrushes. This shift between the end of 2006 and beginning of 2007 is shown in the two histogram figures below for the example of diapers. This is most likely because IRI divides the Academic Dataset in four sets for each period, 2001-2006, 2007, 2008-2011, and 2012, with UPCs reassigned to brands, vendors, types, and category variables between sets. But for the vast majority of months the indexes do not diverge and show no problem with variance or overfitting.



Distribution of Bootstrapped Index Levels End 2006
Diapers
Density
0
50
100
150
200
.87
.875
.88
.885
Index Levels


Distribution of Bootstrapped Index Levels Start 2007
Diapers
Density
0
10
20
30
.7
.75
.8
.85
.9
Index Levels

4 Multilateral Index Results

4.1 CCDI Indexes Over Different Window Lengths

Appendix 2 shows the final CCDI index levels, at month 156, for 24 different window lengths from 2 months, which is a standard monthly chained Törnqvist index, up to 156 months, which is the full span of time, completely circular, and not a rolling window index. Three different imputations methods are shown: (1) the raw, not imputed data; (2) imputing prices for only the nonmissing values; (3) full imputation, for both raw and missing prices. It also shows the same imputation methods for the MPID tests for the same window lengths. All hedonic indexes use a Poisson regression and are weighted by quantity, and the rolling windows are spliced with the half splice method.

If we use the MPID test as the measure of chain drift, all of the raw indexes show significant negative drift that, on average, shrinks with a longer rolling window, until the drift disappears for the full window (by construction). But the path to circularity is far from straight. A common pattern is that after about 3-11 months, the drift grows worse with a longer window, such as for coffee, deodorant, paper towels, salty snacks, toothbrushes, peanut butter and yogurt. It doesn't begin to shrink again until the window length is at least 2 years. For cold cereal, facial tissues, laundry detergent, diapers, razors, and mayonnaise, the drift is still a major problem and doesn't start to improve until after a window of over two years.

The nonmissing imputation levels are usually very close to the raw index levels, except for very short windows, and is actually more likely to be higher than lower. It reduces MPID measured drift for every good for almost every good, improving it in the short windows for all goods except paper towels and coffee, and effectively fixes it entirely for deodorant.

Doing full imputation typically makes a large difference in the index levels. The 2 month chained index levels are usually at least as 'reasonable' as the raw and nonmissing indexes. Aside from a deflation rate of 14% over the 12 years for facial tissues, there is no deflation less than 5% unless there is a divergence as shown in the graphs, which is an indicator that something uncharacteristic happen in those months. Oddly, the full imputation indexes actually have the opposite drift pattern than the raw data for about half of the goods. For 6 of the goods, increasing the window length from 2 months (typical chained) to 9 months makes the chain drift worse, and only for five goods, peanut butter, diapers, deodorant, yogurt, and mayonnaise, does it make an improvement. About half of the goods switch from more(less) drift to less(more) for some range of longer windows. This is very odd, since it would be expected that longer rolling windows should always reduce drift since they do more smoothing. However it should be kept in mind that, as shown in certain months in the bootstrapping results, there could be jumps in some of the full imputation indexes that affect these conclusions.

Overall, moving to a 6 or nine month window makes a big difference for the indexes, but it's only sometimes better and sometimes worse. Moving to an even longer window up to 2 years usually makes little difference, and is also just as likely to have worse drift than better.

The improvement in drift from the hedonic imputation indexes overall, and especially in the 2 month chained indexes, could be because of the smoothing effect hedonic imputation has on prices. The hedonic indexes effectively redefine the goods, so that the consumer has preferences over characteristics instead of over UPCs. The lower drift of the non-missing imputed indexes show the effects of this smoothing without the added complication of the new and exiting items. When there is a sale, only those items on sale have a large price reduction in the raw data. But the hedonic indexes replace those sale prices with predicted prices based on the characteristics coefficients that are estimated from all items, whether on sale or not. Since the sales are not on characteristics themselves, this removes much of the price bouncing that is the underlying cause of chain drift. Unless all items with the same characteristic are on sale in the same month, the predicted price will fall much less than the raw sale price. The fact that some items with that characteristic are on sale will lower that characteristic's coefficient in that month, but the predicted price will still be much less volatile than the raw price. The table shows the correlations between raw and predicted prices for all the categories. The correlations vary from .2454 to .4676, and have a mean of .3554, showing a significant amount of smoothing.

Correlations Between Raw and Predicted Prices

Item	Correlation
cold cereal	.3204
mayonnaise	.4676
coffee	.3915
carbonated beverages	.3004
peanut butter	.4531
razors	.4152
deodorant	.2837
facial tissue	.3861
laundry detergent	.3642
diapers	.3424
tooth brushes	.2739
salty snacks	.2454
paper towels	.3785
yogurt	.3616

5 Conclusions/Discussion

Beyond the obvious reasons for using a hedonic index for the bias reduction of omitting new and exiting goods and avoiding unit values, this paper provides evidence that hedonics can also reduce chain drift, though using full imputation may or may not reduce it. Given the use of a hedonic index, the MPID tests imply that a rolling window multilateral index such as the CCDI is as likely to make drift worse as better unless the window was extremely long. But the odd behavior of the MPID tests could be due to the unreliability of the longer bilateral indexes that make up a window, even a short one, because they are poorer Divisia approximations and suffer from more product turnover. Thus, a better measure of non-circularity should be found before having confidence in any given window length for rolling window multilateral indexes such as the CCDI. Also, if the MPID is unreliable due to the long index lengths, then multilateral indexes with the long windows that supposedly reduce drift are also suspect because they include long index lengths.

6 References

1. von Auer, Ludwig (2024) "Inflation Measurement in the Presence of Stockpiling and Consumption Smoothing". The Review of Income and Wealth, 0, 1-20.

2. Balk, B.M. (2005) "Divisia price and quantity indicies: 80 years after". *Statistica Neerlandica*, 59, 119–158.

3. Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert (1982) "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers". *The Economic Journal*, 92, pp. 73-86.

4. Chessa, Antonio (2025) "Extension of price index series over time for inflation measurement". Unpublished paper.

5. de Haan, Jan and Frances Krsinich. (2014) "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes". *Journal of Business & Economic Statistics,* 32 : pp. 341-358.

6. Diewert, W.E. (1976) "Exact and superlative index numbers". *Journal of Econometrics*, 4, 115–145.

7. Diewert (1993). "The Early History of Price Index Research", in W.E. Diewert and A.O. Nakamura (eds.): *Essays in Index Number Theory, Vol. 1, Contributions to Economic Analysis* 217 (Amsterdam: North-Holland), pp. 33–65.

8. Diewert, W.E. and Kevin Fox. (2022) "Substitution Bias in Multilateral Methods for CPI Construction" *Journal of Business and Economic Statistics*, 40 : pp. 355-369.

9. Divisia, F. (1925) "L'indice monetaire et la theorie de la monnaie. Revue d'economie politique", 39, 842–861.

10. Elteto, O. and P. Koves (1964). On a Problem of Index Number Computation Relating to International Comparisons. Statisztikai 42 (10), 507-518.

11. Gini, C. (1931). On the circular test of index numbers. Metron 9 (9), 3-24.

12. Fox, Kevin, Peter Levell, and Martin O'Connell (2023) "Inflation measurement with high frequency data", Institute for Fiscal Studies Working Paper 23/29.

13. Feenstra, Robert and Matthew D. Shapiro. (2003) "High-Frequency Substitution and the Measurement of Price Indexes" *NBER volume* .

14. ILO. (2004) Consumer price index manual: theory and practice. Geneva: International Labour Organization, International Monetary Fund, Organization for Economic Co-operation and Development, Statistical Office of the European Communities, United Nations, The International Bank for Reconstruction and Development, The World Bank.

15. Ivancic, Lorraine, W. Erwin Diewert, and Kevin J. Fox. (2011) "Scanner data, time aggregation, and the construction of price indexes." *Journal of Econometrics*, 161: pp. 24-35.

16. Hill, R.J. (2006a) "When does chaining reduce the Paasche-Laspeyres spread? An application to scanner data". The Review of Income and Wealth, 52, 309–325.

17. —- (2006b), "Superlative index numbers: not all of them are super". Journal of Econometrics, 130, 25–43.

18. Kurtzon, Gregory. (2022) "How Much Does Formula vs. Chaining Matter for a Cost-of-Living Index? The CPI-U vs. the C-CPI-U" *Economic Inquiry* 60: pp. 645-667.

19. Melser, Daniel and Michael Webster. "Multilateral Methods, Substitution Bias, and Chain Drift: Some Empirical Comparisons". *Review of Income and Wealth*, 67 (2021): pp. 759-785.

20. Pakes, Ariel. "A Reconsideration of Hedonic Price Indexes with an Application to PC's". *The American Economic Review*, 93 (2003): pp. 1578-96.

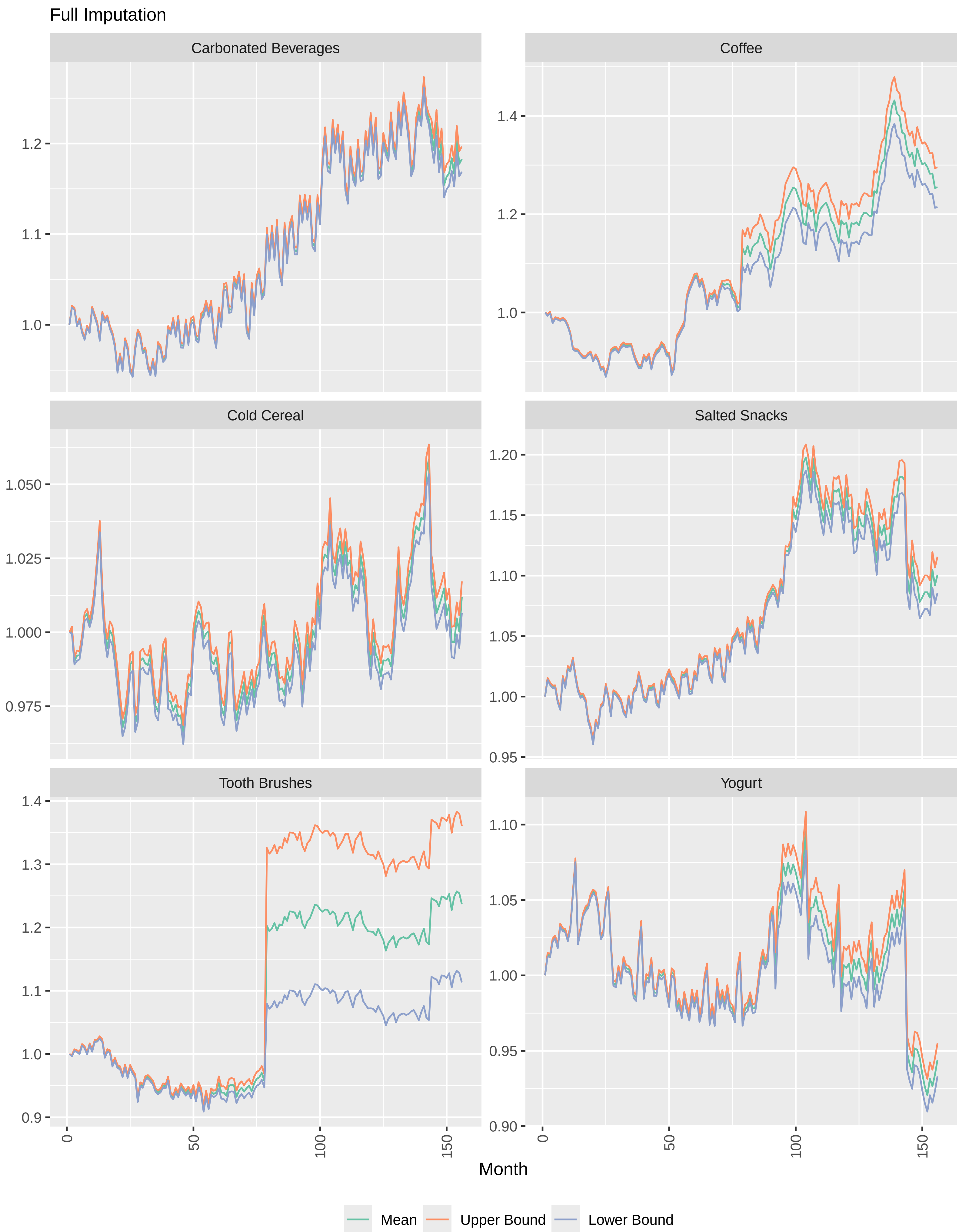
21. Reinsdorf, M.B. (1998) "Divisia indexes and the representative consumer problem". Presented at the Fourth Meeting of the Ottawa Group.

22. Szulc, B. (1964) Indices for Multiregional Comparisons. Przeglad Statystyczny 3, 239-254.

23. Walsh, C.M. (1901) *The Measurement of General Exchange Value.*

7 Appendices

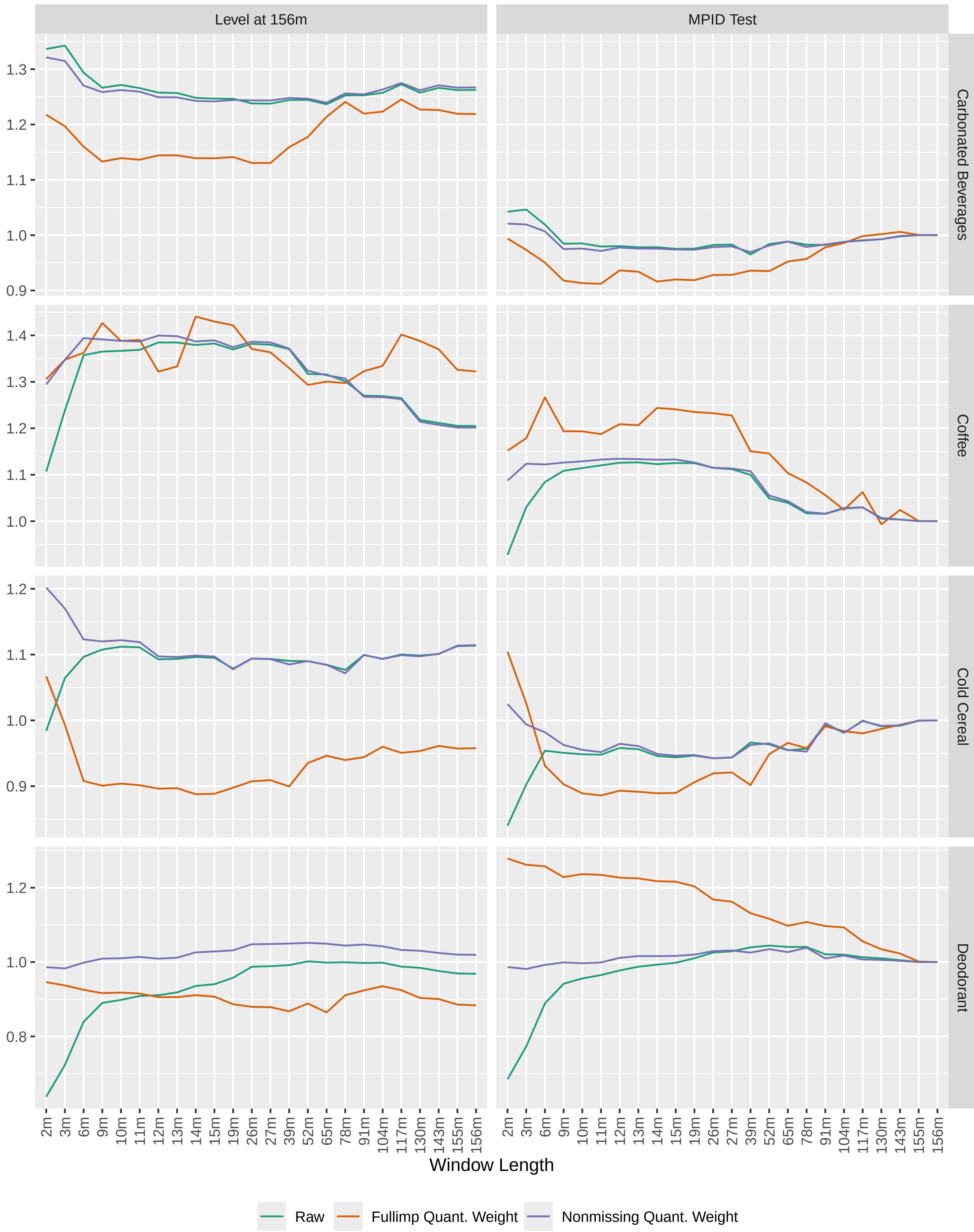
Appendix 1: Mean Bootstrapped Törnqvist Index
Full Imputation Quantity Weighted Poisson

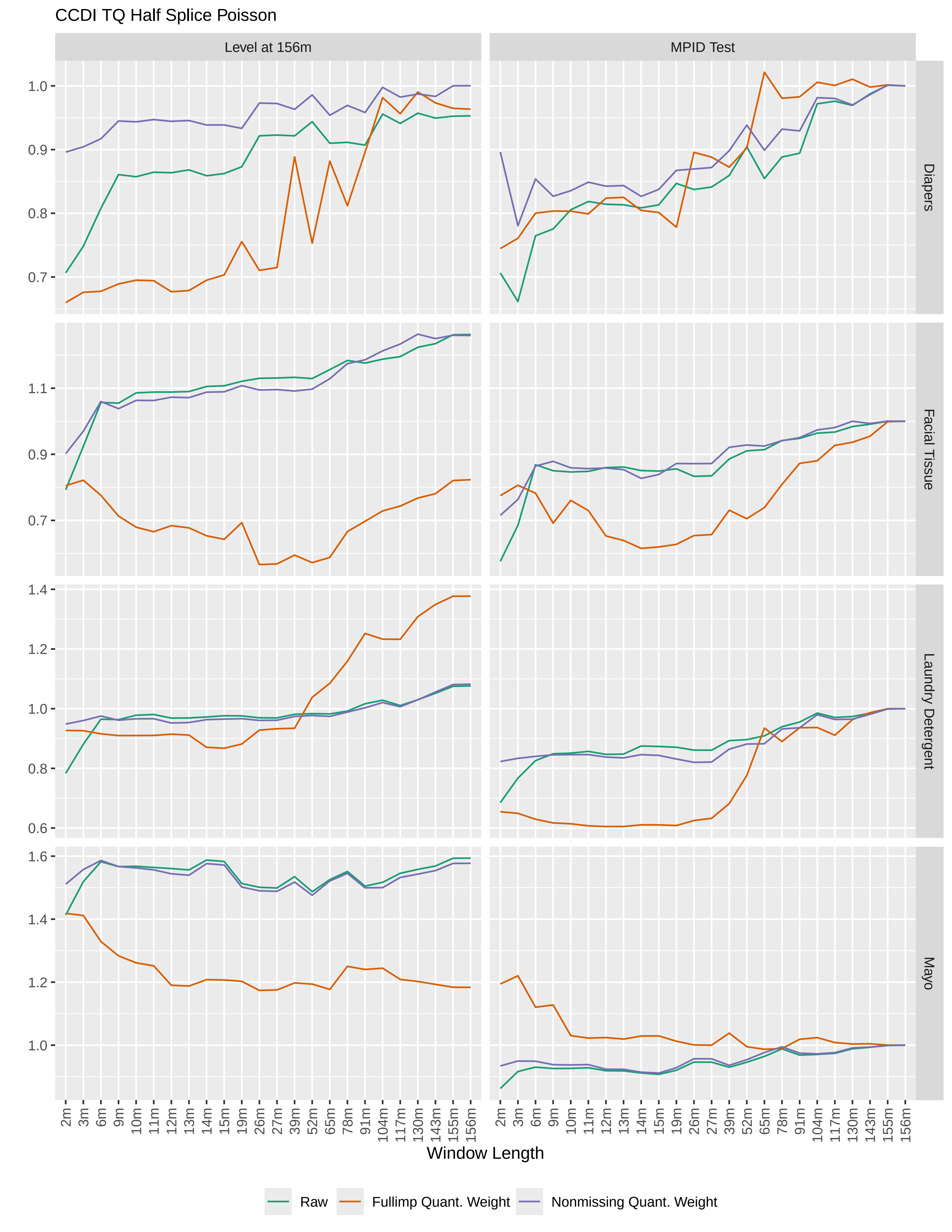


Full Imputation
Carbonated Beverages
Coffee
Cold Cereal
Salted Snacks
Tooth Brushes
Yogurt
Month
Mean
Upper Bound
Lower Bound

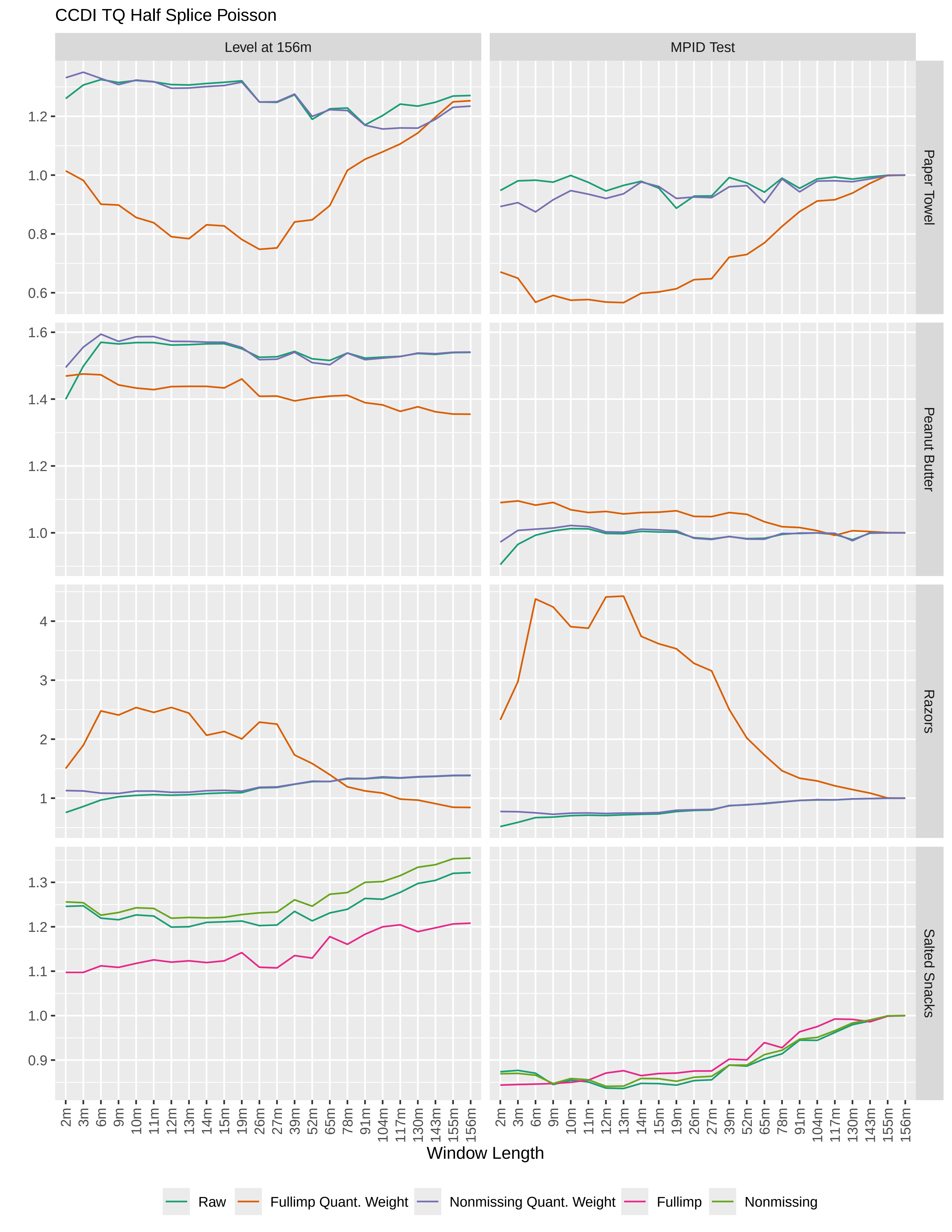
Appendix 2: CCDI Indexes and MPID Tests

CCDI TQ Half Splice Poisson





CCDI TQ Half Splice Poisson
Level at 156m
MPID Test
Diapers
Facial Tissue
Laundry Detergent
Mayo
Window Length
Raw
FullImp Quant. Weight
Nonmissing Quant. Weight

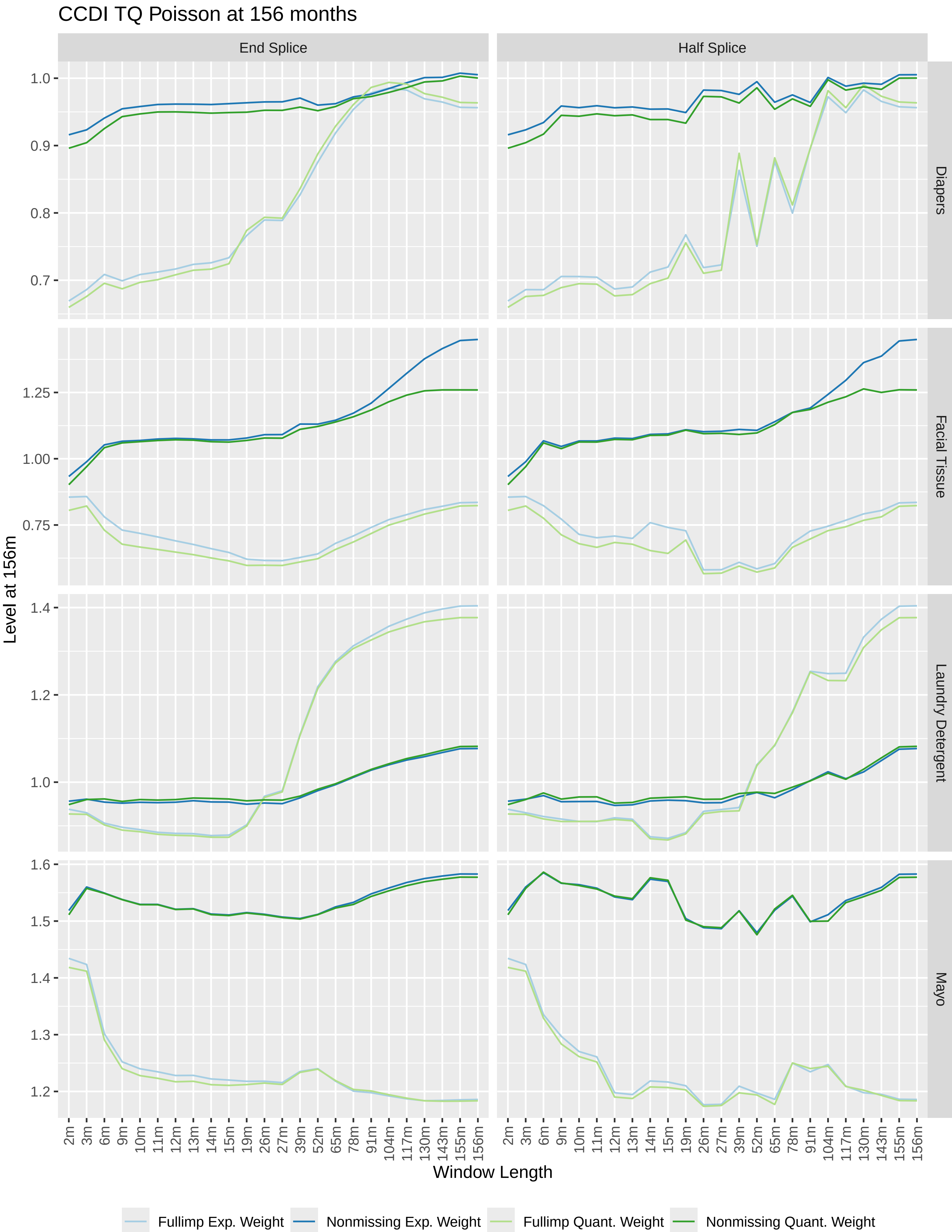


CCDI TQ Half Splice Poisson
Level at 156m
MPID Test
Paper Towel
Peanut Butter
Razors
Salted Snacks
Window Length
2m 3m 6m 9m 10m 11m 12m 13m 14m 15m 19m 26m 27m 39m 52m 65m 78m 91m 104m 117m 130m 143m 155m 156m
Raw
Fullimp Quant. Weight
Nonmissing Quant. Weight
Fullimp
Nonmissing

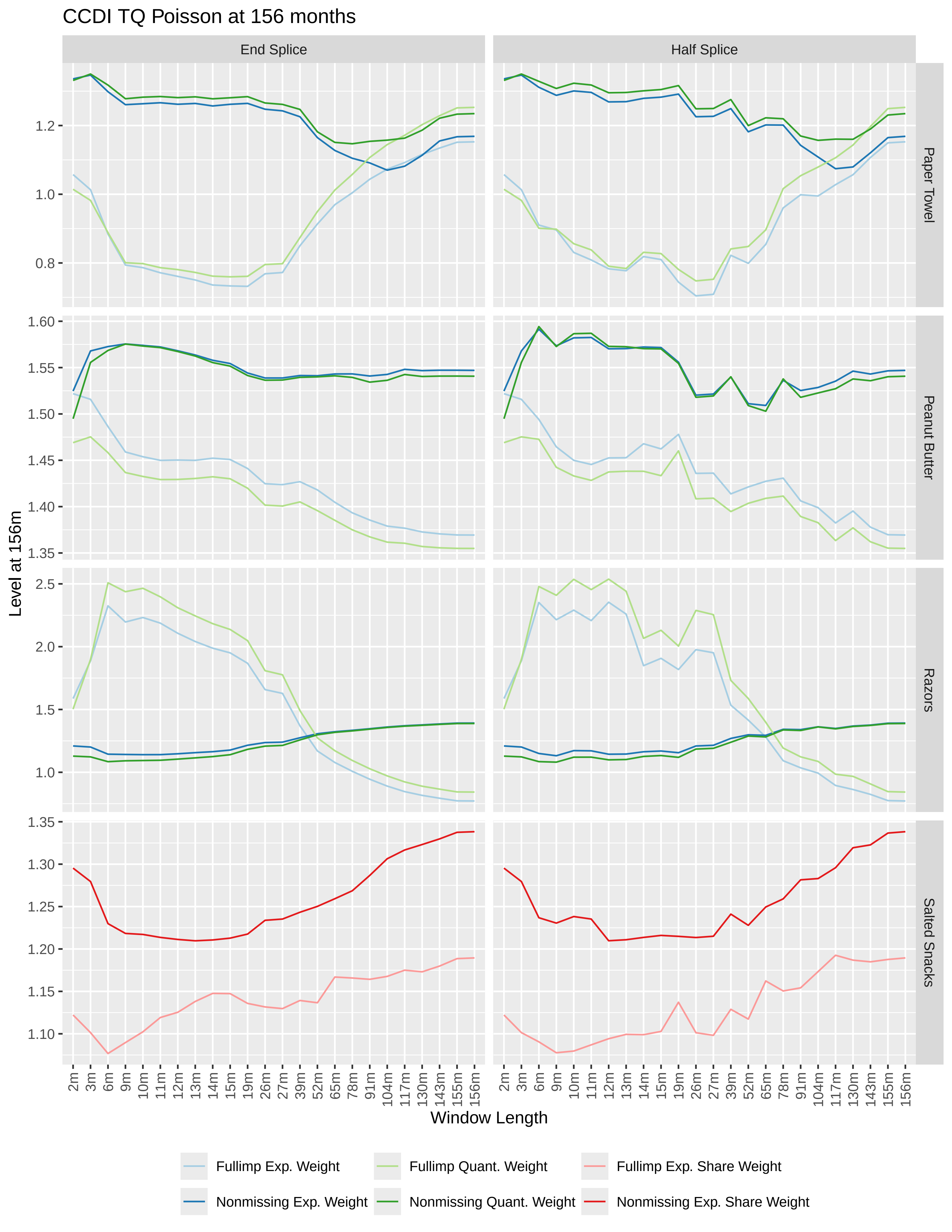
CCDI TQ Half Splice Poisson
Level at 156m
MPID Test
Toothbrushes
Yogurt
1.4
1.2
1.0
0.8
1.2
1.1
1.0
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Window Length
Raw
Fullimp Quant. Weight
Nonmissing Quant. Weight
Fullimp Quant. Share Weight
Nonmissing Quant. Share Weight

Appendix 3: The End Splice vs. Half Splice

CCDI TQ Poisson at 156 months



CCDI TQ Poisson at 156 months
End Splice
Half Splice
Diapers
Facial Tissue
Laundry Detergent
Mayo
Level at 156m
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Fullimp Exp. Weight
Nonmissing Exp. Weight
Fullimp Quant. Weight
Nonmissing Quant. Weight



CCDI TQ Poisson at 156 months
End Splice
Half Splice
Paper Towel
Peanut Butter
Razors
Salted Snacks
Level at 156m
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Fullimp Exp. Weight
Fullimp Quant. Weight
Fullimp Exp. Share Weight
Nonmissing Exp. Weight
Nonmissing Quant. Weight
Nonmissing Exp. Share Weight

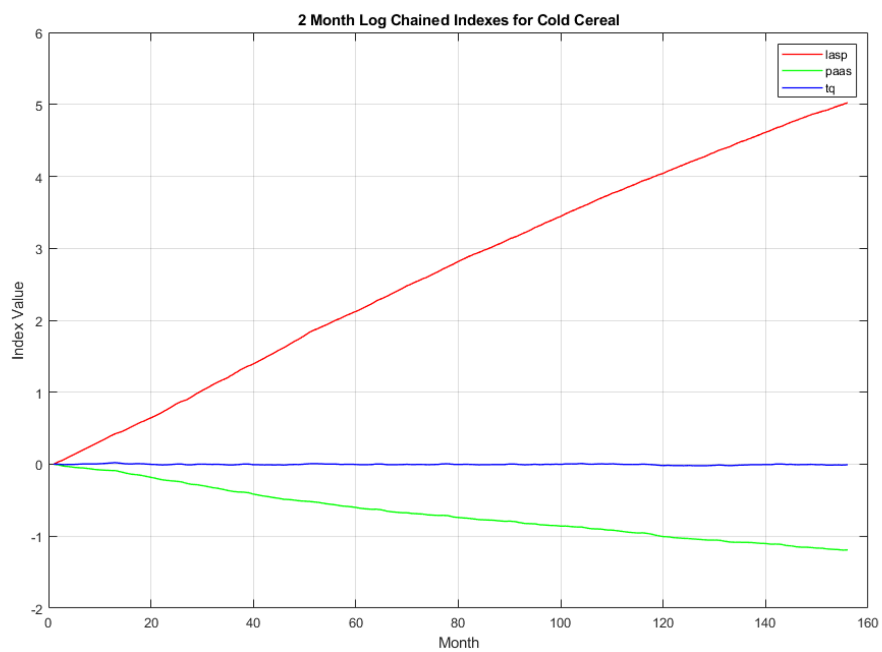

CCDI TQ Poisson at 156 months
End Splice
Half Splice
Toothbrushes
Yogurt
Level at 156m
Window Length
1.75
1.50
1.25
1.00
1.20
1.15
1.10
1.05
1.00
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Fullimp Exp. Weight
Fullimp Quant. Weight
Fullimp Quant. Share Weight
Nonmissing Exp. Weight
Nonmissing Quant. Weight
Nonmissing Quant. Share Weight

4 GEKS Laspeyres-Paasche Spreads

A simpler explanation of the odd results of the MPID tests is that the tests are unreliable. In other words, the MPID test itself, and the multilateral indexes, fail the reasonability test. The longer indexes that make up the windows are less reliable approximations and require more imputations for new and exiting goods than the monthly chained indexes. Even when there is no price bouncing, the direct Törnqvist can diverge from the chained, and the chained is a more reliable index.

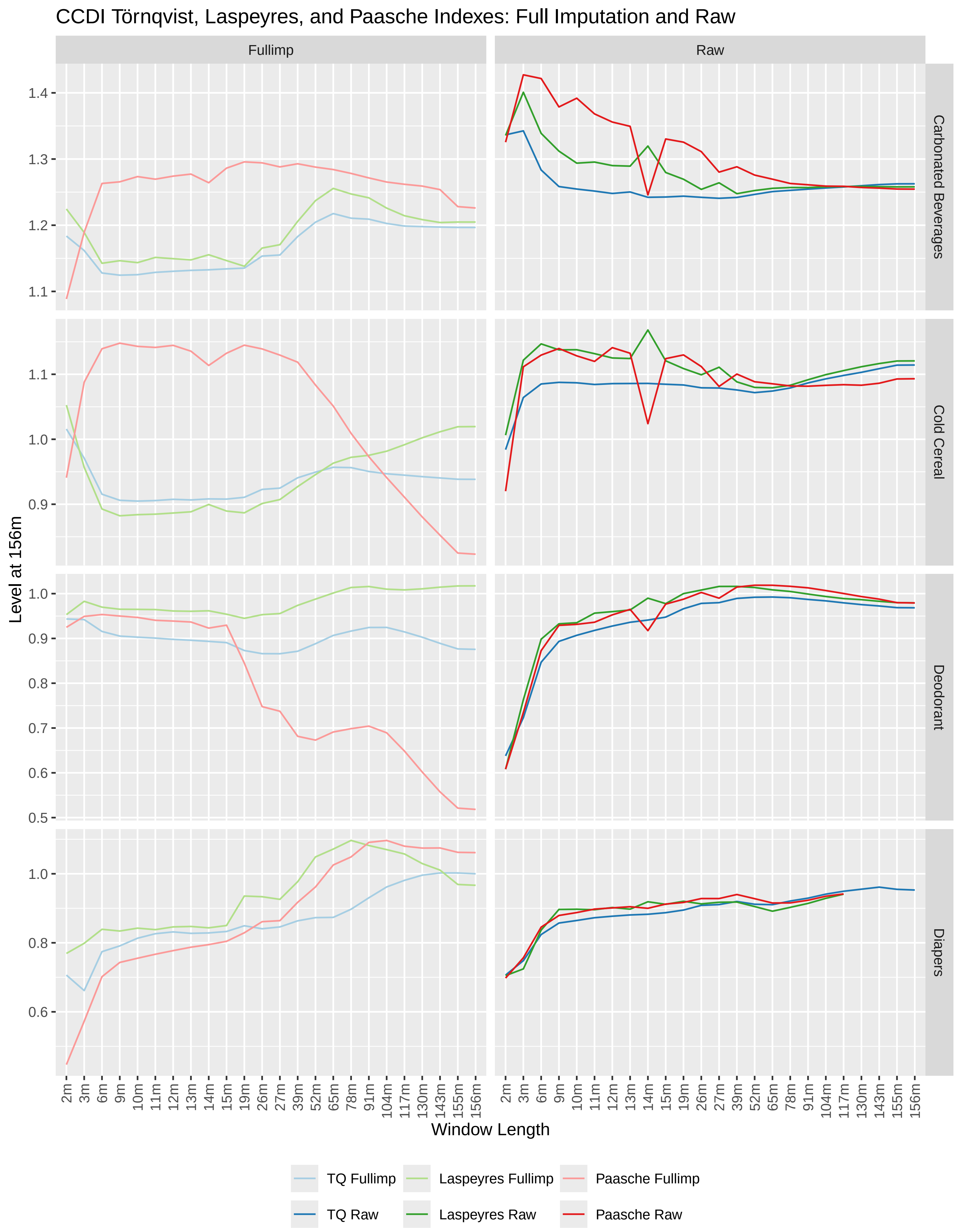
If we can' tuse the full window multilaterals or a full period direct index as the gold standard, we need another estimate of chain drift. As suggested in Hill (2006), since the Laspeyres and Paasche indexes bound a COLI, we can use the Laspeyres-Paasche spread as a measure of drift - a tight spread leaves little room for drift. However, simply using chained Laspeyres and Paasche indexes in this data leaves spreads that are too wide to be useful. For example, the figure below plots the log chained Törnqvist, Laspeyres, and Paasche indexes for cold cereal as an example.

Chained Törnqvist, Laspeyres, and Paasche Indexes for Cold Cereal

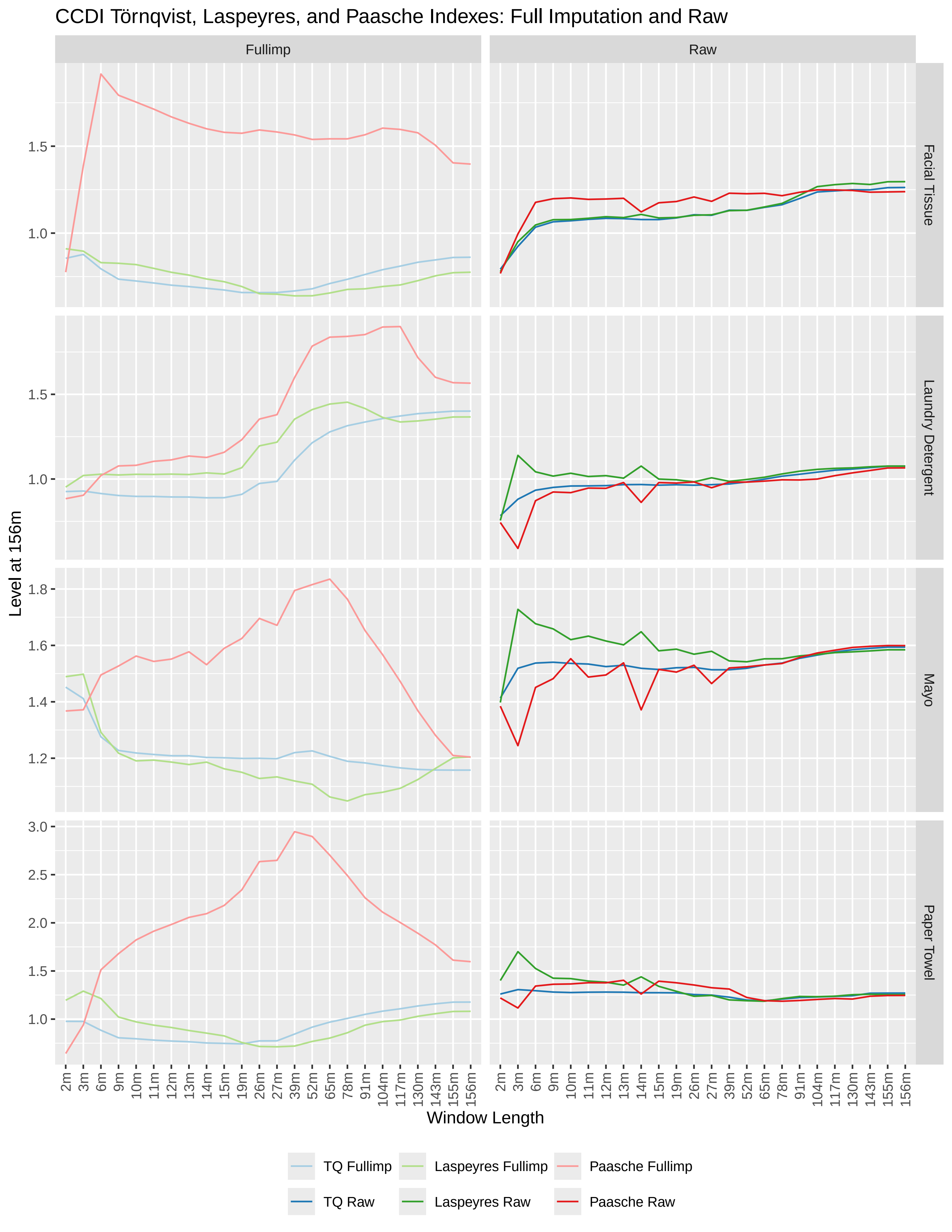


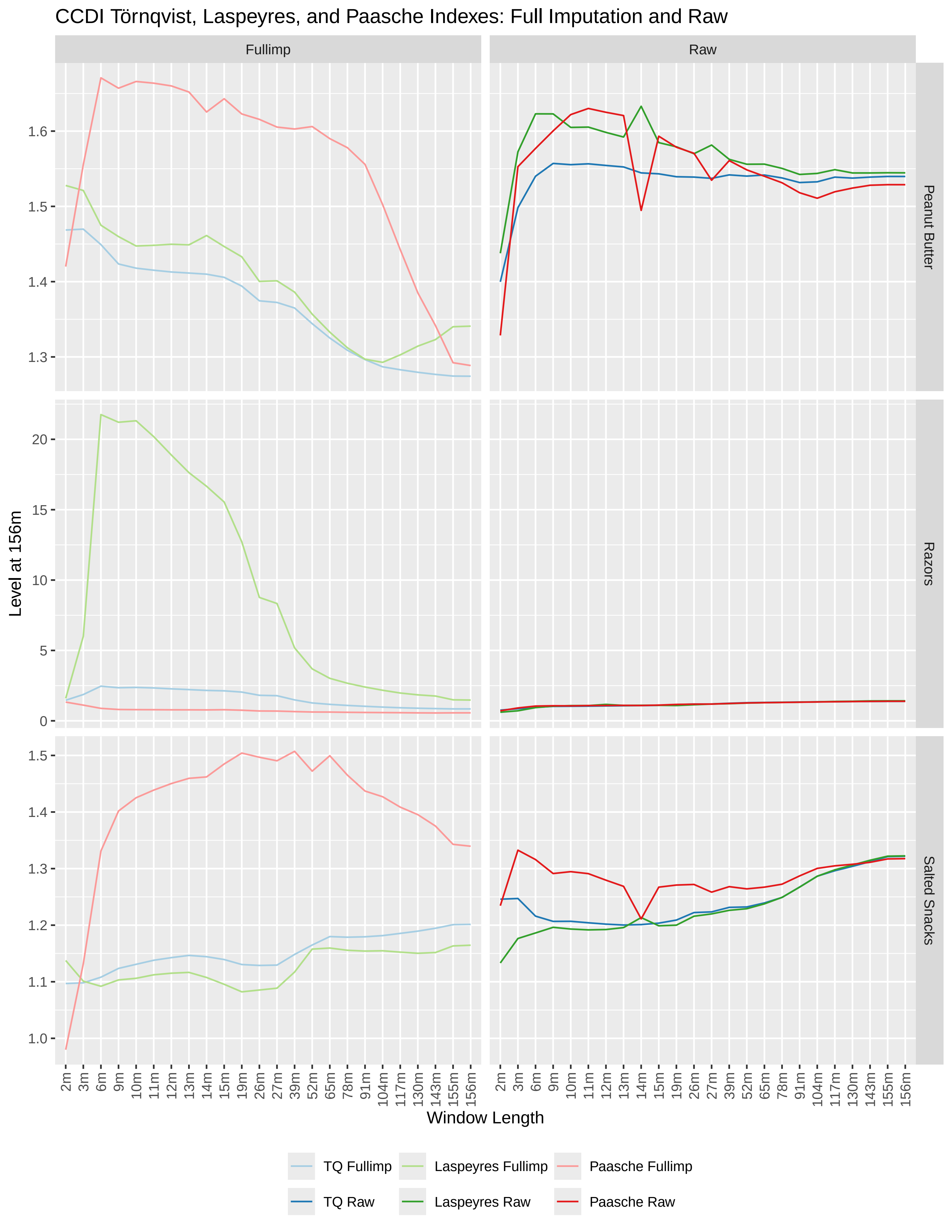
However, we can construct CCDI/GEKS style multilateral Laspeyres and Paasche indexes. Since multilaterals reduce chain drift, the multilateral

Laspeyres-Paasche indexes could provide insight. The 2 month rolling window Laspeyres and Paasche indexes are not the same as the monthly chained version of them because the GEKS procedure averages them with a 1 for the two month window, shown in equation (6). But because the GEKS Laspeyres and Paasche indexes are averages of the bilateral indexes within the window, and the bilateral indexes are bounds on the bilateral COLIs, there may still be insight to be gained. The figures below plot the GEKS Laspeyres and Paasche indexes and the CCDI Törnqvist indexes for the same window lengths as before, for the raw and full imputation hedonic indexes.

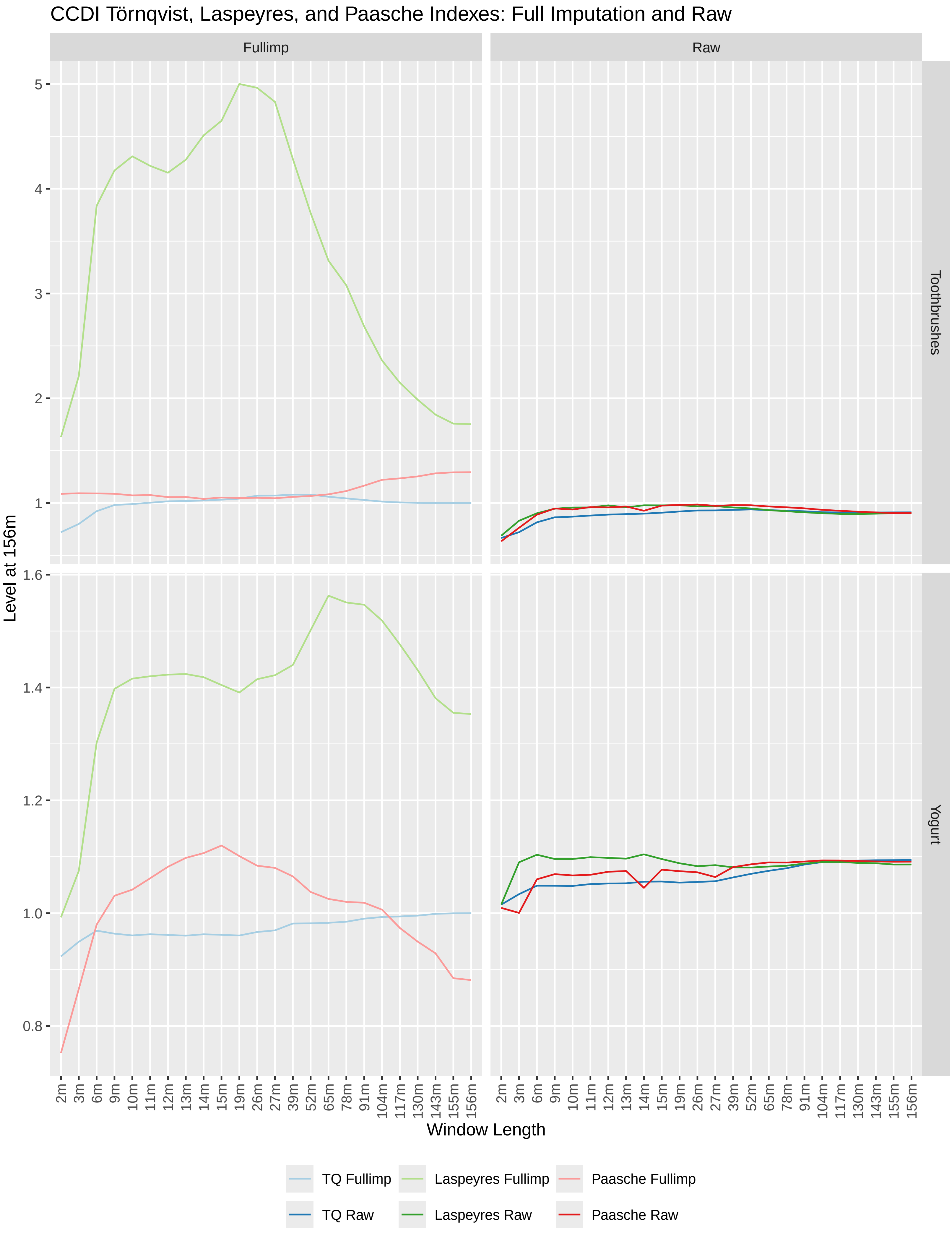


CCDI Törnqvist, Laspeyres, and Paasche Indexes: Full Imputation and Raw
Fullimp
Raw
Carbonated Beverages
Cold Cereal
Deodorant
Diapers
Level at 156m
Window Length
2m 3m 6m 9m 10m 11m 12m 13m 14m 15m 19m 26m 27m 39m 52m 65m 78m 91m 104m 117m 130m 143m 155m 156m
TQ Fullimp
Laspeyres Fullimp
Paasche Fullimp
TQ Raw
Laspeyres Raw
Paasche Raw


CCDI Törnqvist, Laspeyres, and Paasche Indexes: Full Imputation and Raw
Fullimp
Raw
Facial Tissue
Laundry Detergent
Mayo
Paper Towel
Level at 156m
Window Length
2m 3m 6m 9m 10m 11m 12m 13m 14m 15m 19m 26m 27m 39m 52m 65m 78m 91m 104m 117m 130m 143m 155m 156m
TQ Fullimp
Laspeyres Fullimp
Paasche Fullimp
TQ Raw
Laspeyres Raw
Paasche Raw



CCDI Törnqvist, Laspeyres, and Paasche Indexes: Full Imputation and Raw
Fullimp
Raw
Peanut Butter
Razors
Salted Snacks
Level at 156m
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
TQ Fullimp
Laspeyres Fullimp
Paasche Fullimp
TQ Raw
Laspeyres Raw
Paasche Raw



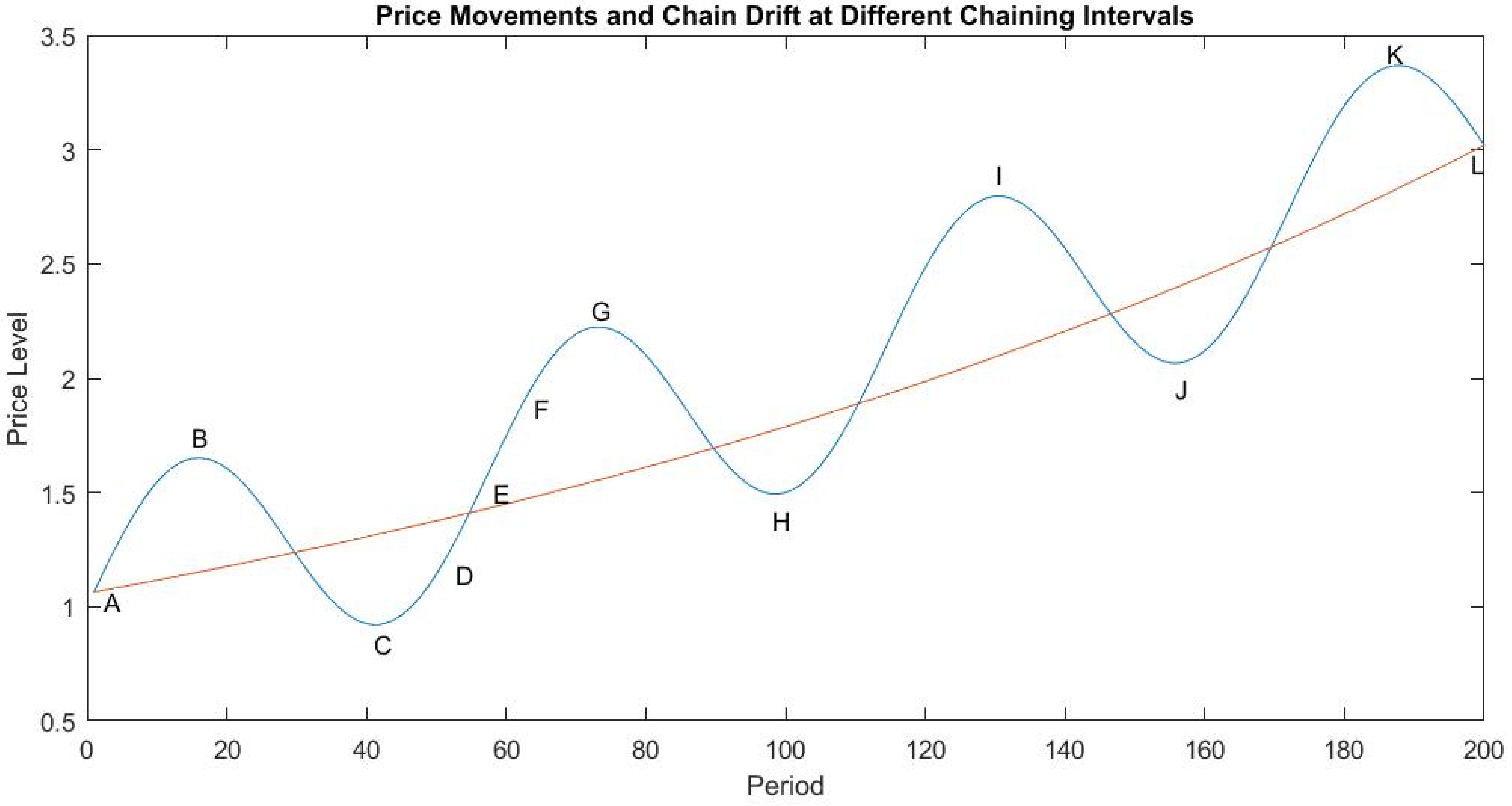
CCDI Törnqvist, Laspeyres, and Paasche Indexes: Full Imputation and Raw
Fullimp
Raw
Toothbrushes
Yogurt
Level at 156m
Window Length
2m 3m 6m 9m 10m 11m 12m 13m 14m 15m 19m 26m 27m 39m 52m 65m 78m 91m 104m 117m 130m 143m 155m 156m
TQ Fullimp
Laspeyres Fullimp
Paasche Fullimp
TQ Raw
Laspeyres Raw
Paasche Raw

In general, the raw spreads are either small for all windows or become tight for longer windows. This is what would be expected, since when there is no chain drift, monthly indexes should all be close approximations of the Divisia index and thus of the true COLI, shown in Kurtzon (2022). Oddly, for short or medium length windows, the Paasche index is above the Törnqvist or even Laspeyres. For the full imputation indexes, the spreads are much wider, often still wide at the full, non-rolling window length. The full imputation GEKS Paasche index is usually above the other indexes even for long windows.

5 The Long Term Törnqvist Index

If the GEKS method itself is the cause of these results, another method for eliminating non-circularity that is more direct to the issue of price bouncing and so uses less irrelevant information than averaging over entire index relatives does. The Long Term Törnqvist (LTTQ) index from Kurtzon (2022) only smoothes out the bouncing in prices.

Figure 3 from Kurtzon (2022) below shows an example of price bouncing that could cause non-circularity.


Price Movements and Chain Drift at Different Chaining Intervals
Price Level
Period
3.5
3
2.5
2
1.5
1
0.5
0
20
40
60
80
100
120
140
160
180
200
A
B
C
D
E
F
G
H
I
J
K
L

Suppose the raw price data for this item is the blue line, from A to B to C to D ... to L. The LTTQ index makes a new smooth price path from A to L without bouncing - the orange line. The long term relative for item i between periods 0 and t, for price p_{it} for item i in period t is

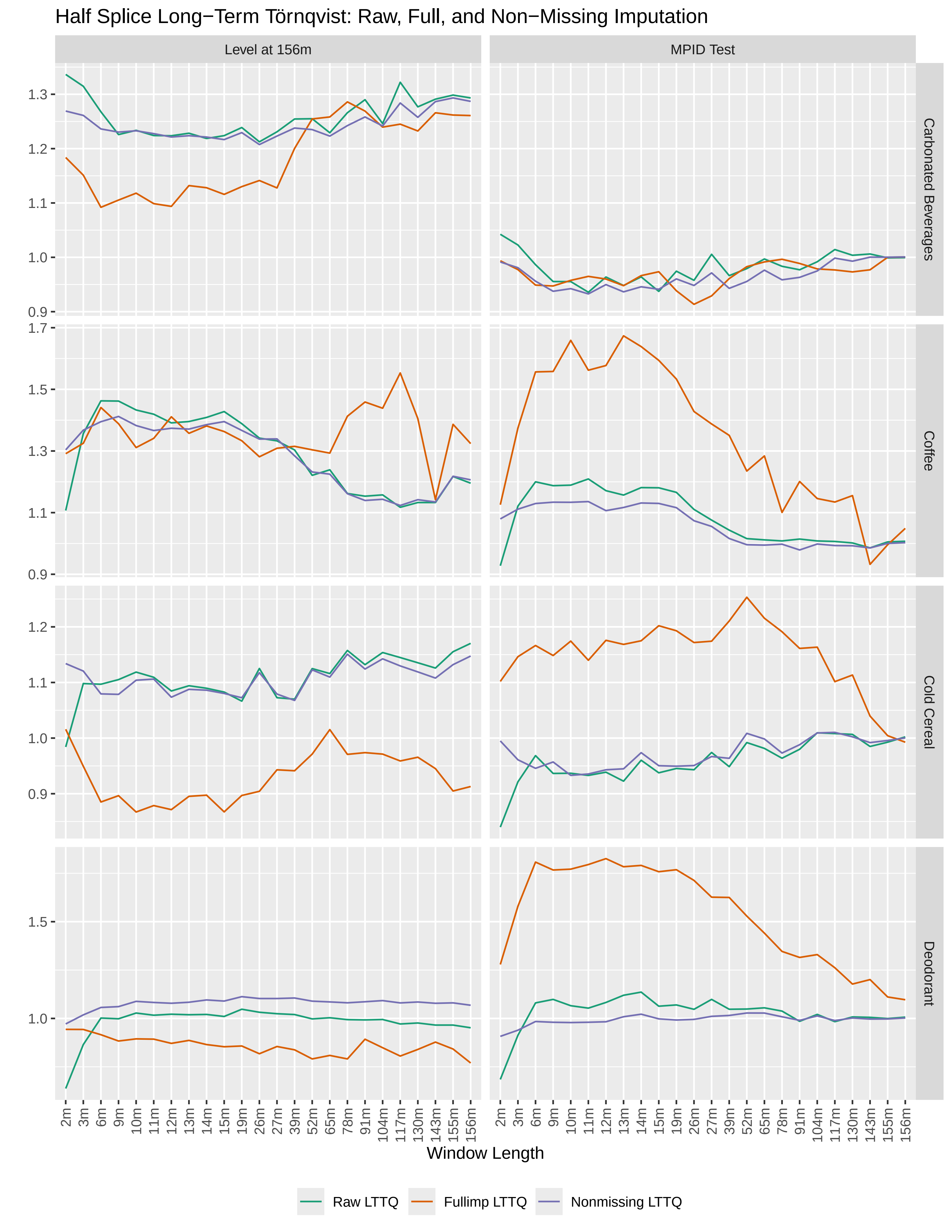
$$d_{it} = \left(\frac{p_{it}}{p_{i0}}\right)^{\frac{1}{t}} \tag{10}$$

. Putting the long term relatives into a Törnqvist formula with T_t^{LT} LTTQ level for period t, is

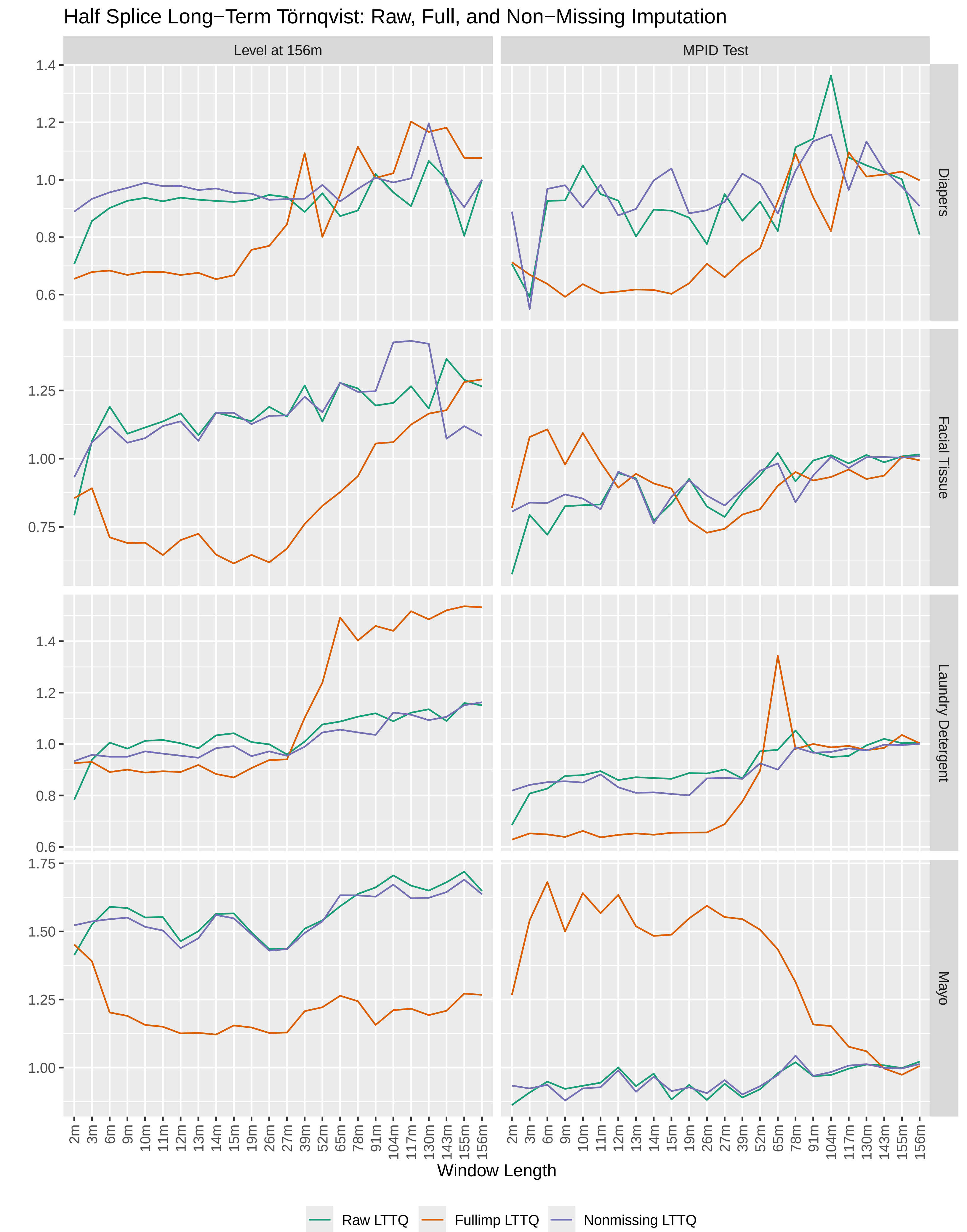
$$\begin{aligned} \frac{T_t^{LT}}{T_{t-1}^{LT}} &= \Pi_i d_{it}^{s_{it}^{TQ}} && (11) \\ &= \Pi_i \left(\frac{p_{it}}{p_{i0}}\right)^{\frac{1}{t}\Sigma_{\tau=1}^{t} s_{i\tau}^{TQ}} && (12) \end{aligned}$$

in terms of the raw prices. In other words, it is a direct Törnqvist index with the average Törnqvist shares in place of the direct Törnqvist shares of $\frac{1}{2}(s_{i0} + s_{it})$.

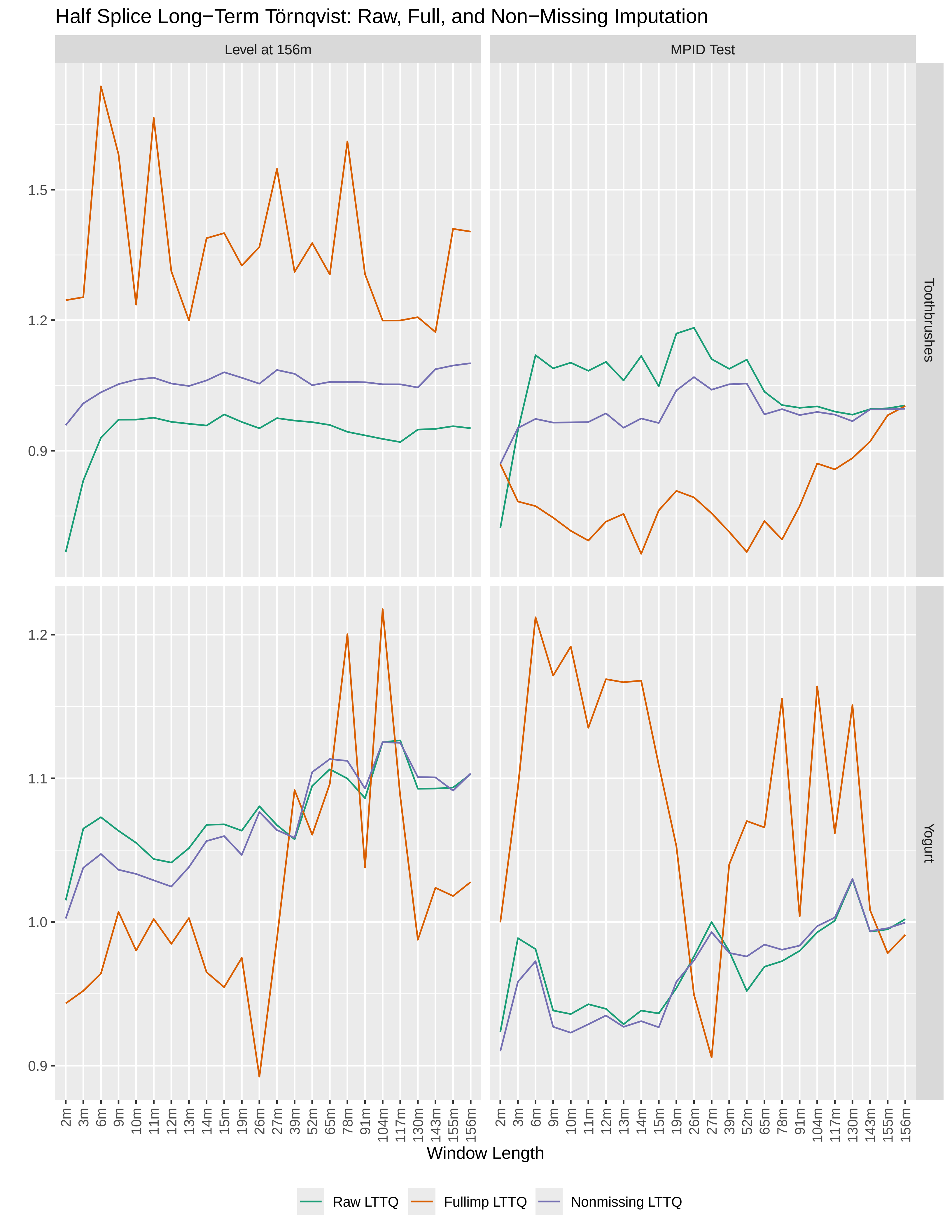
Below are graphs similar to the CCDI results using the LTTQ indexes instead.



Half Splice Long−Term Törnqvist: Raw, Full, and Non−Missing Imputation
Level at 156m
MPID Test
Carbonated Beverages
Coffee
Cold Cereal
Deodorant
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Raw LTTQ
Fullimp LTTQ
Nonmissing LTTQ


Half Splice Long−Term Törnqvist: Raw, Full, and Non−Missing Imputation
Level at 156m
MPID Test
Diapers
Facial Tissue
Laundry Detergent
Mayo
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Raw LTTQ
Fullimp LTTQ
Nonmissing LTTQ

Half Splice Long−Term Törnqvist: Raw, Full, and Non−Missing Imputation
Level at 156m
MPID Test
Paper Towel
Peanut Butter
Razors
Salted Snacks
Window Length
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Raw LTTQ
Fullimp LTTQ
Nonmissing LTTQ


Half Splice Long−Term Törnqvist: Raw, Full, and Non−Missing Imputation
Level at 156m
MPID Test
Toothbrushes
Yogurt
1.5
1.2
0.9
1.2
1.1
1.0
0.9
2m
3m
6m
9m
10m
11m
12m
13m
14m
15m
19m
26m
27m
39m
52m
65m
78m
91m
104m
117m
130m
143m
155m
156m
Window Length
Raw LTTQ
Fullimp LTTQ
Nonmissing LTTQ

In general, they yield similar results, but with slightly higher variance. This could be considered a strength of the CCDI and LTTQ indexes, since they give similar results even using very different methods and smoothing in different ways, or is could be that price bouncing is not the major cause of non-circularity, and something else related to product turnover is.

6 Moving Average Quantities

If non-circularity is caused by stockpiling behavior such as from sales, then one method to eliminate it would be to approximate the correct quantities for a price index to use. The idea is that the quantity that is assumed in how price index formulas are derived is the quantity that consumers would consume if they had to consume everything in the period purchased, living hand-to-mouth. Let this counterfactual quantity be denoted as q^*_{it}. Because consumption can take place over multiple periods and consumers can store goods, these are not the actual consumed quantities, denoted here as q^c_{it}. Let q_{it} denote the transacted quantity in month t. Since over some time period all the goods bought must be consumed, all we really know is that for some window w for item i,

$$\Sigma^t_{t-w+1} q_{it} = \Sigma^t_{t-w+1} q^c_{it} \tag{13}$$

. If consumers pick the actual consumption to be smoothed hand-to-mouth consumption, then

$$\Sigma^t_{t-w+1} q^*_{it} = \Sigma^t_{t-w+1} q^c_{it} \tag{14}$$

so therefore

$$\begin{aligned} \Sigma^t_{t-w+1} q_{it} &= \Sigma^t_{t-w+1} q^*_{it} && (15) \\ \rightarrow & \frac{1}{w}\Sigma^t_{t-w+1} q_{it} = \frac{1}{w}\Sigma^t_{t-w+1} q^*_{it} && (16) \end{aligned}$$

so that mean transacted quantity should approximate the mean hand-to-mouth quantity, and thus approximate the hand-to-mouth quantity of a given period.

Then let the moving average quantity of item i in month t for a window of w months be given by

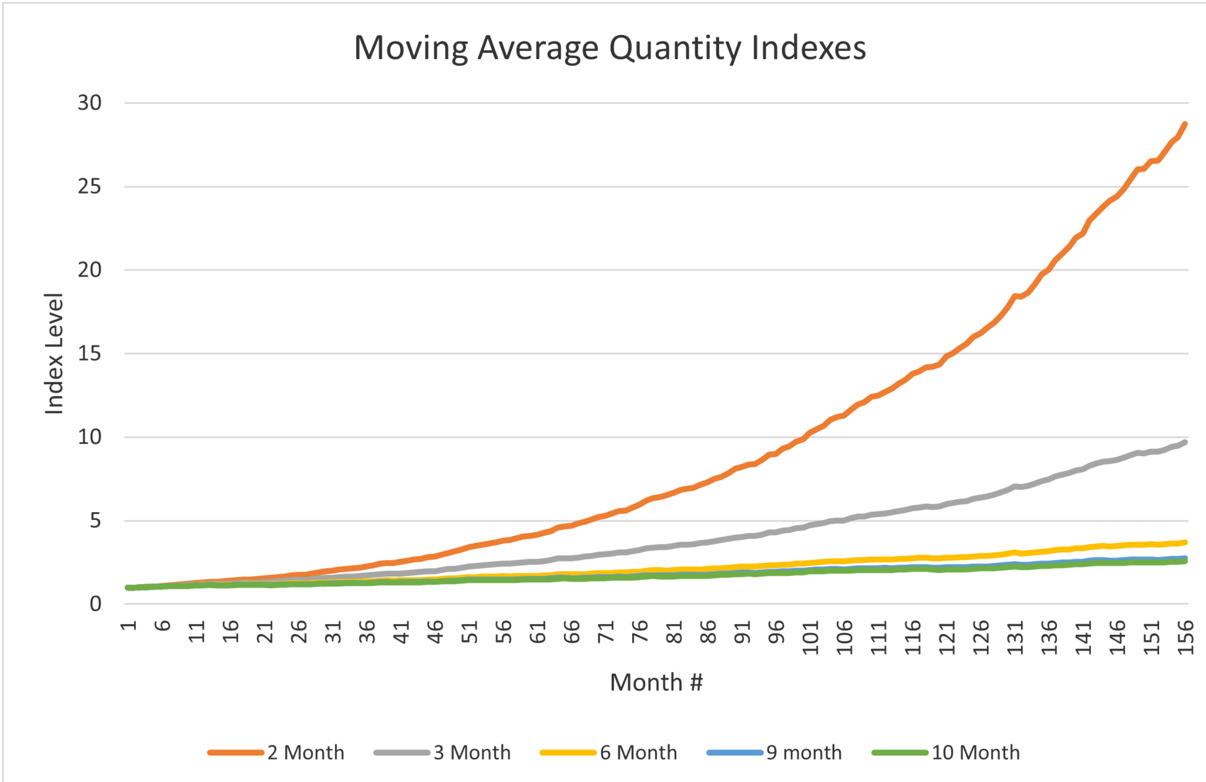
$$q^w_{it} = \frac{1}{w}\Sigma^t_{t-w+1} q_{it} \tag{17}$$

. The moving average quantity expenditure share is then

$$s^w_{it} = \frac{p_{it} q^w_{it}}{\Sigma_j p_{jt} q^w_{jt}} \tag{18}$$

and the Törnqvist share for t-1 to t is

$$s^{TQ,w}_{it} = \frac{1}{2} \quad \frac{p_{i,t-1} q^w_{i,t-1}}{\Sigma_j p_{j,t-1} q^w_{j,t-1}} + \frac{p_{it} q^w_{it}}{\Sigma_j p_{jt} q^w_{jt}} \Bigg) \tag{19}$$



. The figure below shows the moving average quantity indexes for cold cereal, for 2, 3, 6, 9, and 10 month windows. As is obvious, these indexes suffer from even more drift than they are meant to solve, but upward drift instead of downward.

The problem is that the quantities are low in the periods before a sale, so the moving average quantity and thus share is low in the period right before the sale. Then the quantity shoots up during the sale, so the moving average quantities in (17), and thus shares from (18), are very high both during and right after the sale. This means the Törnqvist share from (19) for the downward price movement going on sale, being average of the pre-sale and sale moving average quantity shares, is much lower than the Törnqvist shares right after for the upward price movement going off sale. Longer window lengths dampen but don't eliminate this drift. Perhaps the problem would solved if exactly the right window length for each good was found, so that the moving average would always include one sale quantity in it's window. But many of these goods have so much turnover that they aren't in the sample for more than one sale.