the review of income and wealth

Review of Income and Wealth
Series 59, Number 4, December 2013
DOI: 10.1111/roiw.12068

UNDERSTANDING PRICE VARIATION ACROSS STORES AND SUPERMARKET CHAINS: SOME IMPLICATIONS FOR CPI AGGREGATION METHODS

BY LORRAINE IVANCIC

Centre for Applied Economic Research, University of New South Wales

AND

KEVIN J. FOX*

School of Economics and Centre for Applied Economic Research, University of New South Wales

The empirical literature on price indices consistently finds that aggregation methods have a considerable impact, particularly when scanner data are used. This paper outlines a novel approach to test for the homogeneity of goods and hence for the appropriateness of aggregation. A hedonic regression framework is used to test for item homogeneity across four supermarket chains and across stores within each of these supermarket chains. We find empirical support for the aggregation of prices across stores which belong to the same supermarket chain. Support was also found for the aggregation of prices across three of the four supermarket chains.

JEL Codes: C43, E31

Keywords: aggregation, hedonics, item homogeneity, price indexes, scanner data, unit values

1. INTRODUCTION

Electronic point-of-sale scanner datasets are a relatively new source of data, providing both opportunities and challenges for researchers and price statisticians. Price statisticians now have considerably more access to highly detailed data on consumer purchases—both prices paid and quantities purchased—than ever before. However, research by Reinsdorf (1999), Feenstra and Shapiro (2001), Ivancic *et al.* (2011), and de Haan and van der Grient (2011) has shown that using scanner data to calculate price indexes can lead to highly volatile, and in some cases highly implausible, estimates of price change. This observed volatility seems, in large part, to be directly related to the ability of scanner data to capture frequent, and often large, shifts in quantities purchased in response to changes in price.

This volatility is extremely problematic for researchers trying to model inflation, but perhaps more importantly, for policy makers trying to manage inflation targets. To attempt to stabilize estimates of price change, Reinsdorf (1999) recommended the

Note: The authors gratefully acknowledge constructive comments on a very preliminary draft from participants at the Economic Measurement Group Workshop '07, financial support from the Australian Research Council (LP0347654 and LP0667655), provision of the dataset by the Australian Bureau of Statistics, and very helpful comments from two anonymous referees.

*Correspondence to: Kevin J. Fox, School of Economics and CAER, University of New South Wales, Sydney 2052, Australia (K.Fox@unsw.edu.au).

use of some type of aggregation over quantity and price observations when high frequency data are used. In theory, Reinsdorf's (1999) recommendation to aggregate over prices and quantities seems relatively simple to apply. However, in practice, there are many dimensions (such as time, space, and entity) over which data can potentially be aggregated and, as Hawkes (1997) notes, the choice as to which dimension to aggregate over "is not intuitively obvious"; see Hawkes and Piotrowski (2003) for a more detailed description of potential aggregation units.

The issue of aggregation is fundamental to the construction of price indexes, and hence to estimates of price change. Importantly, different aggregation methods have been shown to impact considerably on estimates of price change, particularly when scanner data are used (Reinsdorf, 1999; Feenstra and Shapiro, 2001; and Ivancic *et al.*, 2011). In addition, choices made about aggregation will reflect different implicit assumptions about consumer behavior. If we choose to aggregate over some unit then we are implicitly assuming that items within that unit are homogeneous, i.e. they are perfect substitutes for one another. As a result, the way we aggregate matters from both a theoretical and a practical perspective. Ideally, good evidence on what are considered to be appropriate methods of aggregation would be available. However, the literature currently provides very little guidance on this issue.

The aim of this paper is to provide a framework within which some basis for making recommendations about aggregation can be made. In particular, we examine whether it is appropriate to aggregate the prices and quantities of an item across different supermarket chains, and across stores which belong to the same supermarket chain. The paper is structured as follows. The issue of aggregation and the property of homogeneity which is associated with "appropriate" aggregation are discussed in Section 2. In Section 3 the definition of homogeneity used in this paper is described. The data and model used in this analysis are described in Sections 4 and 5, respectively. The results of the model are presented in Section 6 and the implications of these results for index number construction are discussed in Section 7. Section 8 concludes.

2. Aggregation, Unit Values, and Homogeneity

In the compilation of a price index, aggregation refers to the calculation of average prices and total quantities *over some unit* such as time, space, or entity. Aggregation over quantities is relatively straightforward. Once the unit to aggregate over has been chosen the quantities relevant to that unit are simply added up. Aggregation over prices involves the construction of a unit value which is, in effect, the calculation of an average price over the aggregation unit.

Ratios of unit values are used by statistical agencies to construct unit value indices, which "are frequently used by countries as surrogates for price changes at the elementary level of aggregation" (Silver, 2010, p. S210). For example, a unit value index between periods 0 and t for commodity group g, UV_g, can be written as follows:

$$UV_g(p^0, p^t, q^0, q^t) \equiv \left(\frac{\sum_{m=1}^{M} p_m^t q_m^t}{\sum_{m=1}^{M} q_m^t} \right) \Big/ \left(\frac{\sum_{n=1}^{N} p_n^0 q_n^0}{\sum_{n=1}^{N} q_n^0} \right)$$

where p_m^t denotes the period t commodity prices for items $m = 1, \ldots, M$ belonging to group g in period t, and q_m^t denotes the corresponding quantities, with period 0 prices and quantities following similar notation for items $n = 1, \ldots, N$ belonging to group g in period 0. If the same set of commodities exists in both periods, then using UV_g to deflate a corresponding index of value change results in an intuitive quantity index, Q_g:

$$Q_g\left(p^0, p^t, q^0, p^t\right) = \left(\frac{\sum_{m=1}^{M} p_m^t q_m^t}{\sum_{m=1}^{M} p_m^0 q_m^0}\right) \Bigg/ \left[\left(\frac{\sum_{m=1}^{M} p_m^t q_m^t}{\sum_{m=1}^{M} q_m^t}\right) \Bigg/ \left(\frac{\sum_{m=1}^{M} p_m^0 q_m^0}{\sum_{m=1}^{M} q_m^0}\right)\right] = \frac{\sum_{m=1}^{M} q_m^t}{\sum_{m=1}^{M} q_m^0}$$

As the numerator and denominator of Q_g involve summing quantities, clearly these have to be homogeneous commodities for this to be appropriate.

From a theoretical perspective, cases exist both for and against the use of unit values in index construction (Diewert, 1995; Balk, 1998; Bradley, 2005). However, from a purely practical perspective, unit values may provide price statisticians with a method to maximize the use of the price and quantity information contained in scanner datasets while minimizing the associated problems of index number volatility. Finding a method to facilitate the maximum use of information contained in scanner datasets for index number construction is the primary motivation of this research.

It is generally accepted that aggregation should occur across units which are homogeneous or alike (Dalen, 1992; Reinsdorf, 1994; Balk, 1998). Silver (1999, 2009, 2010) shows that not aggregating when goods are in fact homogeneous leads to index numbers which are biased (and aggregating when goods are heterogeneous results in unit value indexes which are biased). Considering the case of "broadly comparable" items, he follows the work of de Haan (2004a, 2007) in recommending quality-adjusted unit value indexes; this approach adjusts for quality differences using a hedonic model before aggregating over the items. That is, an attempt is made to make the items more homogeneous before aggregation.

The issue of when a group of items, stores, or time periods are thought to be sufficiently homogeneous to justify aggregation leads to some difficult questions. For example, we need to consider how narrow or broad a category should be and what characteristics—e.g., across items, stores, and time—are considered to be important in determining "sameness." These issues lead to the following question which underlies the appropriate construction of unit value, as stated by Balk (1998, p. 1):

> [W]hen is a commodity (group)—that is, a set of economic transactions—sufficiently "homogeneous" to warrant the use of unit values?

Balk (1998, p. 9) showed that if "the unit value index is appropriate for a certain commodity group then it is equal to each single price ratio, and all those price ratios are equal. Thus, the observation of only one commodity suffices to calculate the price index." Balk (1998) also noted that in practice "small distortions" in price may occur. In this paper we propose the use of a hedonic model to test for item homogeneity while taking account of these small distortions in price.

There are many units over which we can potentially aggregate and hence, many units over which homogeneity can potentially be tested. The focus of this paper is to determine whether we can provide some information on homogeneity in the following contexts:

1. If the same item is found in different stores which belong to the same supermarket chain, should we consider the item to be homogeneous across stores within a supermarket chain?
2. If the same item is found in different supermarket chains, should we consider the item to be homogeneous across supermarket chains?

3. Testing for Homogeneity

As supermarkets tend to compete on price rather than quantity we assume a Bertrand competition framework.[1] In Bertrand competition, where there are two or more firms which sell a homogeneous item the standard outcome is that price will equal marginal cost, i.e. we will obtain a competitive market outcome. If any seller in this market chooses to undercut their competitors' price they will be able to capture the entire market (Spulber, 1999). However, with product differentiation this outcome will not hold, with firms able to maintain higher (lower) prices than their competitors without losing (capturing) the entire market.

In practice, different supermarket chains stock a large number of items which are physically identical. We argue that if price dispersion is found to exist in this market then it is due to "product" differentiation—where the differentiation is not embodied in the physical product itself but in the range or quality of services offered by different retailers. This may include different opening hours and differences in the provision of customer parking and sales staff. Store location, and in particular, store accessibility (or convenience) is also considered to be a service characteristic (Betancourt and Gautschi, 1988). If consumers value these services the same item may be sold at a higher price by sellers who offer a relatively higher level of complementary services. In such a market the price of an item reflects a bundle of both the item and seller attributes. In this case, Reinsdorf (1992) noted that "apparent retail market price dispersion would be an artefact of measuring only part of the bundle priced by retailers." Therefore, we treat the level of auxiliary services provided by a seller as important in determining when a good is homogeneous and as a result, when it is appropriate to construct unit values.

The Bertrand framework—where price differences for the same good can exist when there is product differentiation—underlies the definition of homogeneity used in this paper, which is as follows.

> *Definition*: The same item sold by different sellers is viewed as homogeneous if the price of the item is found to be consistently the same across sellers in the long term.

[1]Two supermarket pricing strategies defined in the marketing literature are Every Day Low Pricing (EDLP) or Hi-Lo pricing, where temporary price discounts are important.

An issue to consider is that of imperfect or costly consumer information.[2] It may be argued that this type of framework relies on all consumers knowing all prices offered by all sellers at any point in time and that, in practice, this is not the case. We do not dispute this point. However, we assume that if *persistent* price differences exist, consumers learn about these price differences over time and in the long run the consumer will move their purchases to the seller that gives them the best price–service combination.

4. Data Description

This study uses an Australian scanner dataset containing 65 weeks of data, collected between February 1997 and April 1998. It contains information on 110 stores which belong to four supermarket chains located in the metropolitan area of one of the major capital cites in Australia. These stores accounted for over 80 percent of grocery sales in this city during this period (Jain and Abello, 2001).

Data on the item category “coffee” were used, due to the standard yet ubiquitous nature of multiple versions of the product. Two supermarket chains dominate the market and together account for approximately 75 percent of the expenditure for this particular item category. The dataset includes information on all instant coffee items sold in all stores. Information on each item includes the average weekly price paid for each item in each store in each week, the total quantity of that item sold in each store in each week, a short product description (including information on brand name, product type, flavor, and weight) and a unique numeric identifier for each item. The unique identifier allows for the exact matching of items over time. In total, the dataset has 514,945 weekly observations on 205 items across all stores. A number of data exclusions were made, which consisted primarily of items which were not thought to belong directly to the coffee item category, such as “coffee substitute” cereal beverages. Additional data deletions were also made due to “missing” information such as store, weight, or brand. The excluded items accounted for 5.4 percent of total expenditure in this item category. After data exclusions 436,103 weekly observations on 157 coffee items remained.

The dataset identifies the store in which the item was sold and the supermarket chain to which the store belongs. The identification of supermarket chains in such datasets appears to be quite rare, with the authors knowing of only one other study (Bradley *et al.*, 1997) to have used such information. However, different chains are identifiable only by number, not by name. Information on each supermarket chain, including the number or stores and coffee items sold, is shown in Table 1.

The four supermarket chains vary considerably in size, both in terms of their market share and in terms of the number of stores which comprise the chain. The smallest chain accounts for only 4 percent of the total expenditure in this item category and is comprised of nine stores. In comparison the largest chain accounts for 41 percent of the total expenditure in this item category and has 41 stores which

[2]Different transport costs for chains should not be issue as they are all located in the same capital city, both in our empirical work and in terms of typical consumer choice.

TABLE 1
Descriptive Statistics

	No. Stores	Expenditure Share	No. Coffee Items Sold in Each Chain	No. Weekly Observations	No. Monthly Observations
		%			
Chain A	26	20	89	89,320	22,381
Chain B	9	4	101	29,155	8,063
Chain C	34	35	123	162,765	41,853
Chain D	41	41	88	154,863	38,953
Total	110	100	157	436,103	111,250

belong to this chain. The detailed information in the dataset on the store and supermarket chain in which an item is found, along with the description of the coffee items, makes it possible to estimate a hedonic model to test for item homogeneity across stores and chains.

5. The Hedonic Model

A hedonic regression model—where the price of an item is regressed on the characteristics of that item—is used to test for potential homogeneity across sellers. Two widely used alternative approaches for specifying the hedonic model are the time dummy variable (TDV) method and the exact hedonic approach. The general form of the TDV hedonic model is defined by Silver and Heravi (2007) as:

$$\ln p_i^t = \beta_0 + \sum_{t=2}^{T} \beta^t D^t + \sum_{k=1}^{K} \beta_k Z_{ki}^t + \varepsilon_i^t, \tag{1}$$

where p_i^t is the price of item i in period t, D^t is a time dummy variable for periods $2, \ldots, T$, and Z_{ki} denotes the set of K characteristics of item i, where $k = 1, \ldots, K$.

In the standard TDV approach the B_k coefficients, representing the value of the characteristics, are fixed across time and each observation is given an equal weighting. In the exact approach the characteristics are allowed to vary and weights are explicitly incorporated into the model (Silver and Heravi, 2003). In our model we use elements of both approaches. As the data cover a relatively short period (15 months) and we have no reason to believe that significant changes occurred in the values of the characteristics coefficients in this period, we restrict the coefficients of the characteristics to be constant over time (following the TDV approach). However, we consider weighting issues to be important, particularly with our dataset, so weights were explicitly incorporated into the model; the issue is the relative importance of an item.

5.1. *The Economic Importance of an Item*

Accounting for the economic importance of an item in a market is considered extremely important in the index number literature. Typical weights considered in this type of analysis include quantity of sales, total item expenditure or, equivalently, expenditure share. Diewert (2003) notes that the use of quantity weights "will tend to give too little weight to models that have high prices and too much weight to cheap models that have low amounts of useful characteristics." Results are reported for models where expenditure shares (i.e., the expenditure share of item *i* in period *t*) were used to weight the observations. This weighting approach obviously amounts to using weighted least squares (WLS) estimation.

As WLS is typically used to correct for heteroscedasticity it is of interest to test whether the error terms were heteroscedastic, and if so, whether the use of WLS would correct for this heteroscedasticity.[3] White's (1980) test found heteroscedasticity to be present in the unweighted model and the WLS model. As a result, standard errors were corrected for heteroscedasticity using the procedure of White (1980).

5.2. *Package Weight and Spline Functions*

A feature of many items sold is that the price of the item is often not linearly related to the weight of the item. Fox and Melser (2012) examined the issue of non-linear pricing using scanner data for the item category "soft drinks." The authors found significant discounts available for larger package sizes and multi-packs and that the actual relationship between prices and volumes was significantly flatter than that indicated by a linear relationship. Diewert (2003) recommended that such non-linear pricing can, and should, be captured in a hedonic regression. In particular, Diewert recommended the use of a piecewise linear, continuous spline function to capture this relationship.

Twenty different coffee jar weights, ranging from 49.6 grams to 1000 grams were identified. The weights in the range of 49.9 grams to 125 grams were combined into the first category and the weights ranging from 375 grams to 1000 grams were combined to make up the last weight category. This was to avoid the well-known end-point problem.[4] This left us with seven potential knots in our spline function. These potential knots were located at the following weights (in grams): 126, 151, 201, 251, 301, 341, and 376. As there was little *a priori* evidence to determine where the knots should be located, standard model selection criteria including the Akaike Information Criteria (AIC), the Bayesian Information Criterion (BIC), and the adjusted R-squared were used to select the best model.[5] Both the AIC and BIC were minimized when five knots were included in the model at the following weights: 126, 151, 201, 251, and 341 grams. The adjusted R-squared was also highest for this model.

5.3. *Model Variables*

A number of product characteristics for the range of coffee items were identified from the dataset. Characteristics included in the model were:

[3]We also wanted to ensure that if errors were homoscedastic in the unweighted model then the use of weighting would not introduce heteroscedasticity.

[4]"As there are fewer data points constraining the fit near the end points of the approximation, there is the possibility that the spline function may fit the data very closely in these areas, resulting in low bias but unacceptably high variance" (Fox, 1998). For an early use of splines in the economics literature, and linear splines to overcome the end point problem, see Fuller (1969).

[5]Specifications of AIC and BIC are found in Appendix A.

- *Product brand*: Dummy variables were specified for each brand. In total, 25 different brands were identified.[6]
- *Decaffeinated product*: A dummy variable was used to indicate whether the product was decaffeinated.
- *Additional flavoring*: A dummy variable was used to indicate whether the coffee product had any additional flavoring such as chocolate, vanilla, amaretto etc.
- *Product weight*: Twenty different package weights for the coffee items were identified, ranging from 49.6 grams to 1000 grams. A number of items were excluded as the item weight was not recorded in grams. Product weight entered into the model as a spline function.
- *Bonus*: A dummy variable was used to indicate whether a product contained a bonus, e.g. more weight at a reduced price.

Additional model variables include:

- *Time*: A time dummy variable was included for each month after the first. Although weekly data were available, the data were aggregated to generate monthly observations. This was done as it was thought that if weekly data were used, price differences across chains may simply reflect differences in the timing of sales across chains rather than meaningful differences in price. Aggregating the observations over each month should, to some extent, smooth out the effects of different timing of sales across chains.
- *Supermarket chain*: Information on four supermarket chains was available. Dummy variables were created for each chain.

In this study, the supermarket chain in which an item is sold is viewed as simply another characteristic of the coffee item. Therefore, the supermarket chain variable enters the TDV hedonic regression model of equation (1) via the Z_{ki} item characteristics variable. The WLS estimating equation, which is equivalent to weighting observations by expenditure shares, is then as follows:

$$(2) \qquad \sqrt{\exp_{ic}^{t}}\ln p_{ic}^{t} = \sqrt{\exp_{ic}^{t}}\beta_0 + \sum_{t=2}^{T}\sqrt{\exp_{ic}^{t}}\beta^{t}D^{t} + \sum_{k=1}^{K}\sqrt{\exp_{ic}^{t}}\beta_k Z_{kic}^{t} + \sum_{c=2}^{C}\sqrt{\exp_{ic}^{t}}\beta_c H_{ic} + \sqrt{\exp_{ic}^{t}}\varepsilon_{ic}^{t}$$

where p_{ic}^{t} is the price of item i, in supermarket chain c, in period t, D^t is a dummy variable for time periods, $t = 2, \ldots, T$, Z_{kic} is the set of K characteristics, $k = 1, \ldots, K$, of item i in chain c, H_{ic} is a dummy variable for supermarket chain, $c = 2, \ldots, C_i$, where item i is sold, ε_{ic}^{t} is a stochastic error term, and the expenditure share of item i, in supermarket chain c in period t is denoted by:

$$\exp_{ic}^{t} = \frac{p_{ic}^{t}q_{ic}^{t}}{\sum_{i=1}^{I}\sum_{c=1}^{C_i}p_{ic}^{t}q_{ic}^{t}}.$$

[6]A referee notes that with chain-specific brands, using both chain dummies and dummies for brands could lead to multicollinearity problems. However, store brands comprise only 3.51 percent of total expenditure over our period, with some differences between chains: Chain A: 7.17%, Chain B: 2.01%, Chain C: 2.08%, Chain D: 3.15%. There was no empirical evidence of multicollinearity.

TABLE 2
Coffee Expenditure Market Share for Each Chain

	Chain A	Chain B	Chain C	Chain D
Month 1	0.187	0.044	0.327	0.408
Month 2	0.190	0.042	0.342	0.395
Month 3	0.191	0.038	0.343	0.396
Month 4	0.187	0.034	0.332	0.418
Month 5	0.206	0.037	0.338	0.391
Month 6	0.212	0.034	0.323	0.403
Month 7	0.209	0.036	0.326	0.399
Month 8	0.186	0.034	0.339	0.413
Month 9	0.201	0.041	0.332	0.393
Month 10	0.179	0.041	0.347	0.375
Month 11	0.194	0.034	0.315	0.401
Month 12	0.174	0.043	0.323	0.400
Month 13	0.183	0.043	0.334	0.374
Month 14	0.220	0.039	0.321	0.363
Month 15	0.172	0.041	0.320	0.398
Min	0.172	0.034	0.315	0.363
Max	0.220	0.044	0.347	0.418

The β_c coefficients on the supermarket chain dummy variables, H_{ic}, are our coefficients of interest. Note that we are assuming a common coefficient across the items for each supermarket chain. Thus, we are not directly testing the homogeneity of a single item across chains, but rather the (weighted) average homogeneity over all coffee items. Put another way, we are testing for quality (of service) differences reflected in prices across chains, assuming that the quality differences are common to all coffee items.[7] The significance or insignificance of these chain dummy coefficients will tell us which chains, if any, it is appropriate to construct unit values over. We now turn to the results.

6. Results

The aim is to determine whether the same item found in different supermarket chains can be considered as homogeneous. This can be answered by examining if *persistent* differences in prices for the same items exist across chains; as argued in Section 3, price differences between sellers offering the same level of service should lead to the relatively higher priced sellers losing market share as consumers move to the relatively lower priced sellers. In our dataset the expenditure shares of the four supermarket chains are very stable across time; see Table 2. This shows that if persistent price differences are found across supermarket chains, these differences indicate different levels of service or quality across the chains.

6.1. *Hedonic Regression Results*

In our initial hedonic regression analysis, unit values are constructed over the same item across stores in each chain. This reflects the assumption that the same

[7]While this may be a reasonable assumption for an item category such as coffee, this may not be such a reasonable assumption for other item categories, such as clothing.

TABLE 3

Results for Price Differences Across Chains: Aggregation Over Stores Within a Chain

	Chain A	Chain B	Chain C	Chain D
Chain A	–	0.0307	0.0374	0.0325
		(0.115)	(0.053)	(0.115)
Chain B	0.0307	–	0.0067	0.0017
	(0.115)		(0.711)	(0.931)
Chain C	0.0374	0.0067	–	–0.0050
	(0.053)	(0.711)		(0.8005)
Chain D	0.0325	0.0017	–0.0050	–
	(0.115)	(0.931)	(0.8005)	

Note: Coefficients are reported with the corresponding p-values for the chi-square statistics in parentheses.

item sold in different stores which belong to the same supermarket chain are considered to be homogeneous. This assumption seems to be broadly in line with statistical agency practice. For example, the Australian Bureau of Statistics notes the following:

> Large retail chains frequently have a common Australia-wide or state-wide pricing policy. In these cases, pricing one outlet in the chain would be considered sufficient to obtain a representative estimate of price movement for that chain. (ABS, 2005, p. 75)

The ABS (2005) goes on to say that in practice "the usual procedure is to have a number of observations in the samples commensurate with their overall market shares."

The overall fit of the hedonic regression model appears to be quite good, with an adjusted R-squared of 0.69. In general, the signs and magnitudes of most of the coefficients appear to be reasonable. The primary interest of the model is to determine whether the estimated coefficients on the supermarket chain variables are significantly different from one another. For ease of interpreting the differences between pairs of chains, the model of Section 5 was run four times, with each supermarket chain used as the "base" chain to obtain the associated coefficients and p-values between all pairs of chains. In Table 3, the first column indicates which supermarket chain was used as the "base" chain in the regression model. The table includes the estimated coefficient for each of the supermarket chain variables, with the associated chi-square statistic p-values in brackets. In the first row of Table 3 we test whether coffee item prices in Chain B, C, and/or D are significantly different from prices in the base chain, Chain A. Full model results are presented in Appendix B, Table B1.

The results show that after adjusting for the various coffee types included in each supermarket chain, (log) coffee item prices are not found to be significantly different across the supermarket chains, at a strict 5% significance level. However, we can see that the hypothesis of no price differences across Chains A and C is only just not rejected ($p = 0.053$) and the evidence for no price differences between Chains A and B, and A and D are not particularly strong (both

TABLE 4
Results for Price Differences Across Chains: No Aggregation Over Stores Within a Chain

	Chain A	Chain B	Chain C	Chain D
Chain A	–	0.0313*	0.0343*	0.0326*
		(<0.01)	(<0.01)	(<0.01)
Chain B	0.0313*	–	0.0030	0.0013
	(<0.01)		(0.585)	(0.821)
Chain C	0.0343*	0.0030	–	–0.0017
	(<0.01)	(0.585)		(0.615)
Chain D	0.0326*	0.0013	–0.0017	–
	(<0.01)	(0.821)	(0.615)	

Note: Coefficients are reported with the corresponding p-values for the chi-square statistics in parentheses.
*Indicates significance at the 1% level.

p = 0.115).[8] These results are based on the assumption that all stores within a supermarket chain have similar pricing policies (i.e., homogeneity across stores within a supermarket chain) and as such, unit value prices were constructed across all stores. It is of interest to test whether this restriction is in fact appropriate.

To test the assumption of (weighted) average homogeneity of coffee items across stores within a supermarket chain the regression equations were run again, but this time unit values were not constructed across stores within the same chain for the same item. That is, unit values were constructed for each item in each store. For this model, observations included the average price paid and the total quantities purchased of each item in *each* store. In total, there are 111,250 observations. Again, standard model selection criteria were used to choose the "best" number of knots in the weight spline. The "best" model was found to have all seven weight knots included. The findings of this analysis are presented in Table 4, with the full model results presented in Appendix B, Table B2. The first column in Table 4 indicates which supermarket chain was used as the "base" chain in the regression model.

The results presented in Table 4 appear to be fairly consistent with those shown in Table 3 in terms of the sign and magnitudes of the store dummy variable coefficients. The main difference is that they seem to provide much stronger support for the existence of significant price differences between Chain A and all other chains. At the 1% significance level the hypothesis that prices in Chain A are the same as those in Chains B, C, and D, respectively, is rejected.

The difference in these results suggests that the method of aggregation and the significance level chosen may matter in terms of when we find the assumption of homogeneity to be satisfied. Based on this uncertainty, it seems prudent to explore the issue of whether it is sufficient to assume homogeneity across stores located in a particular chain more thoroughly. Therefore, the next issue of interest is to determine which, if any, supermarket chains allow different pricing policies across stores and whether we can make any generalizations about constructing unit values at the level of stores that belong to a particular supermarket chain.

[8]Note that as we have a semilogarithmic model with dummy variables, the (β) coefficients in the table can be transformed to allow interpretation as percentage differences in price levels between chains following Halvorsen and Palmquist (1980): $100\{\exp(\beta) - 1\}$. See also Goldberger (1968) and Kennedy (1981). This implies that prices levels in chains B, C, and D are approximately 3–3.8 percent higher than in Chain A.

6.2. *Testing for Store Differences Within a Chain*

The model used in this section is similar to the model outlined above except for some minor variations. First, separate regression equations are run for each supermarket chain. A variable which indicates which store in a chain an item belongs to was included in each regression equation. Within each equation we were then able to test whether prices in stores which belong to the same supermarket chain differ significantly after controlling for the characteristics of different coffee items. For each chain regression equation the AIC, BIC, and adjusted R-squared were again used to determine the number of knots in the weight spline. If all three model selection criteria did not give the same result, then the model which was selected by two out of the three criteria was chosen.

The results of the pair-wise store comparisons show very little price variation across stores that belong to Chain A and Chain B.[9] In Chain A no pair of stores out of 325 comparisons was found to have significant price differences. In Chain B, no significant price differences were found across any pair of stores. Chain C shows some minor variation in prices, with prices in eight pairs of stores (or approximately 1.4 percent) found to be significantly different from each other. In Chain D the results show slightly more variation, with 61 of the 861 pair-wise store price comparisons (or approximately 7.0 percent) found to be significantly different from each other. If we take a closer look at the price variation in Chain D, two stores seem to have consistently different prices from most other stores in Chain D and it is these stores which account for the majority of price variation found. If these two stores (which we label stores 1 and 2) are excluded from our analysis, only 2 of the 861 (or approximately 0.2 percent) pair-wise store comparisons are found to be significantly different from each other.

Although there is no formal test or cut-off point to determine homogeneity, based on these results it appears reasonable to say that the same coffee item sold in stores belonging to Chains A and B appears to be homogeneous as no stores within these chains were found to have significantly different prices. In Chain C, as the price variation across stores was negligible, the assumption of homogeneity also appears reasonable. As homogeneity appears to be satisfied for these three supermarket chains, the use of unit values across stores within these chains seems reasonable. For Chain D, it seems that the assumption of homogeneity is largely satisfied for all stores except stores 1 and 2. Therefore, constructing unit values across all stores in Chain D, except stores 1 and 2, also seems reasonable.

These results inform us about aggregation across stores which belong to the same supermarket chain. However, a higher level of aggregation may be possible. Based on the results in Table 3 we can also establish whether aggregation across supermarket chains is reasonable. There is fairly clear evidence that prices in Chains B and C, B and D, and C and D are not significantly different from each other. However, the evidence is not as clear cut for price differences in Chains A and B, A and C, and A and D. Therefore, aggregation of Chain A data with chains B, C, and D is not recommended. As the ultimate purpose of this analysis is to inform us

[9]Full results are available from the authors on request.

about appropriate index number construction, it is of interest to compare index number estimates using the aggregation methods suggested by the above analysis.

7. Index Number Estimation using Different Aggregation Methods

The impact of different aggregation methods on four well known indexes—the Laspeyres, Paasche, Fisher, and Törnqvist—is explored in this section; see Appendix A for the respective index number formulae. The Laspeyres and Paasche indexes do not account for consumer substitution over time. The Fisher and Törnqvist indexes are known as superlative indexes (Diewert, 1976) and have been shown to provide a second order approximation to a Cost of Living Index (COLI).[10] That is, these indexes provide an approximation to an index which can account for consumer substitution. It is of interest to see whether the impact of different aggregation methods is different across superlative and non-superlative indexes.

The different aggregation methods to be compared include the following:

- *Homogeneity over all chains*: The same good is considered to be homogeneous no matter which store or supermarket chain it is found in. Unit values are constructed over the same item across all stores.
- *No homogeneity over chains*: The same good is not considered to be homogeneous over chains but is considered to be homogeneous over stores within a supermarket chain. Unit values are constructed over the same item found within stores located in the same supermarket chain.
- *No homogeneity across stores*: If the same item is located in different stores it is treated as heterogeneous. No unit values constructed across stores or chains.
- *Homogeneity based on hedonics*: The same item is considered to be homogeneous if it is found in stores which formed part of Chain A. Unit values were then constructed across the same item for stores located in Chain A. The same good is considered to be homogeneous if it is found in stores which form part of Chains B, C, and D—except for stores 1 and 2 located in Chain D. Unit values were then constructed across the same item found in Chains B, C, and D (except for stores 1 and 2). Stores 1 and 2 entered index number calculation separately.

Chained (updated base) indexes are calculated for each of the different aggregation methods described above. Table B3 of Appendix B presents month-to-month estimates of price change and a chained estimate of total price change over the 15-month period, organized by index number formula. Figure 1 presents the same information, but organized by aggregation method.

As expected, higher levels of aggregation tend to lead to more stable estimates of price change over the 15-month period. It can be seen that index number estimates based on hedonics give us a considerable level of aggregation. However, the impact of using different aggregation methods does not appear to be consistent across the different index number formulae. The indexes which do not, by

[10]For a more detailed explanation on the Cost of Living Index, see chapters 17 and 18 of the CPI manual (ILO, 2004).

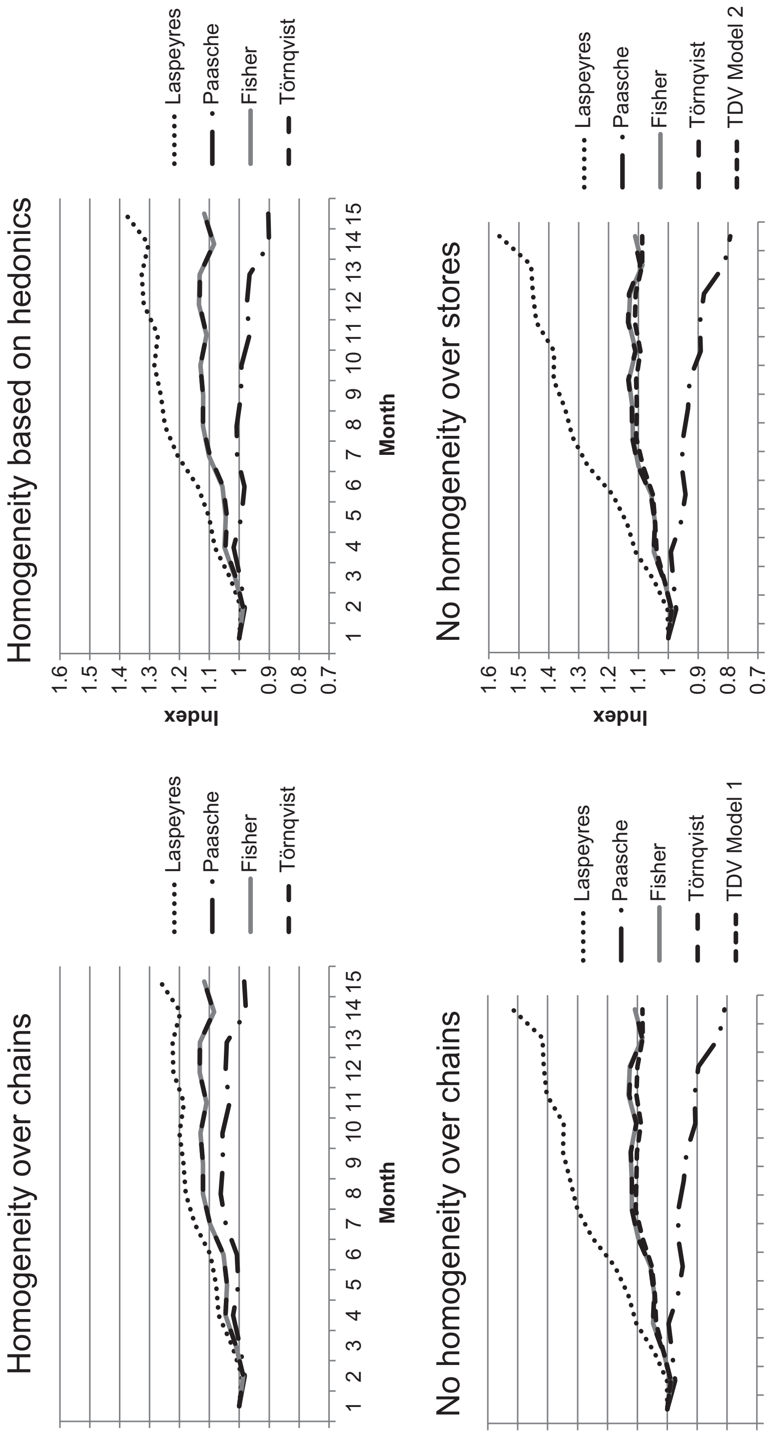

Figure 1. Index Number Estimates Using Different Types of Aggregation

construction, allow for consumer substitution (i.e., the Laspeyres and Paasche) are most noticeably affected by the different aggregation methods. For example, from Table B3 the different aggregation methods used with the Laspeyres index lead to index number estimates of price change ranging from 26 to 56 percent. As can be seen clearly from Figure 1, the indexes which can allow for some consumer substitution appear to be much less affected by aggregation, price change estimates for the superlative indexes ranging from approximately 11 to 12 percent.[11]

The results show that when non-superlative index numbers are used to calculate price change, aggregation choices can have a huge impact. However, the issue of aggregation seems to become relatively trivial when the standard Fisher and Törnqvist superlative indexes are used, with an extremely close range of estimates of price change found across different aggregation methods.[12] This result seems to provide further support for the use of these superlative indexes over the use of non-superlative indexes to estimate price change. In general, this study shows that, when non-superlative index numbers are used, it seems to be particularly important to try and test in some ways the underlying aggregation assumptions that are made when calculating the price index, as the assumptions may have a considerable impact on the final estimates of price change.[13]

There are two further indexes which we can plot by way of comparison. In Section 6, the time dummy variable (TDV) hedonic regression equation (2) was estimated under two alternative sets of assumptions; no homogeneity over chains (unit values were constructed across stores within the same chain for the same item), and no homogeneity over stores (unit values were constructed for each item in each store). From the coefficient estimates in Tables B1 and B2, respectively, we can construct TDV price indexes, by taking the exponent of the coefficients on the time dummies and continuously compounding.[14] The results are plotted in the lower two panels of Figure 1; TDV Model 1 uses the coefficients from Table B1, and TDV Model 2 uses the coefficients from Table B2. The TDV price indexes are unchained indexes which are "quality adjusted" as they account for new and disappearing items, whereas the other indexes are strictly matched-item indexes without quality adjustment. From Figure 1, the two TDV indexes are very similar to the superlative indexes, lying just below them in each case.[15] That is, the

[11]The similarity between the Fisher and Törnqvist series under each method of aggregation, such that the series are effectively overlapping in Figure 1, indicates that we would be empirically indifferent about which of these indexes to choose.

[12]The use of non-standard superlative indexes can lead to a spread which is bigger than that between Paasche and Laspeyres indexes; see Hill (2006).

[13]The use of chained superlative indexes may not be without problems when using scanner data. It particular, there may be chain drift. See Ivancic *et al.* (2011) for a solution to this problem, the Rolling Window GEKS approach. Note that they use the same dataset as in this study, yet their results for coffee (their tables 2–7) are slightly different to ours. This is due to data omissions in the current study given the information requirements for the alternative aggregation approaches being considered; see Section 4. They found the amount of drift for coffee to be negligible. See de Haan and Grient (2011) for more on the approach of Ivancic *et al.*

[14]The adjustment of Goldberger (1968) for the coefficients of dummy variables in semilogarithmic models makes no significant difference to the results.

[15]de Haan (2004b) showed that for a particular choice of weights in the WLS regression, a time dummy price index can be interpreted as a Törnqvist index, where the prices for new and disappearing goods are imputed from the regression results. The share weights used here for the WLS regression do not lead to this interpretation.

quality-unadjusted superlative indexes slightly overstate price change over the period relative to these the TDV indexes.[16]

A further issue of interest is to understand whether estimates of price change were similar across supermarket chains. In theory, if all supermarket chains have the same *rate* of price change, then sampling items from one supermarket chain only would provide a "representative" estimate of price change across items in all chains. Estimates of price change, using the Laspeyres, Paasche, Fisher, and Törnqvist indexes, for each of the four supermarket chains are presented in Table B4 of Appendix B. Coffee items were not matched across supermarket chains for these indexes.

Table B4 includes month-to-month estimates of price change, a chained estimate of price change, labeled "total," for the 15-month period and the percentage change in the final column. Figure 2 presents a graphical representation organized by supermarket chain. The results indicate that for index number formulas which allow for consumer substitution—i.e., Fisher and Törnqvist indexes—estimates of price change are very similar across chains, with the difference between the highest and lowest rate of "total" inflation estimated at 5.8 and 6.1 percent, respectively. Although Figure 2 reveals that the series for Chain A are slightly more volatile than for other chains, the total rate of price change in chains A and B appears to be particularly close when the Fisher or Törnqvist indexes are used. These types of similarities do not appear when indexes which cannot account for consumer substitution, such as the Laspeyres and Paasche indexes, are used. Differences between the lowest and highest rates of inflation across the supermarket chains are estimated at 47 and 21 percent, respectively, for the Laspeyres and Paasche indexes. From the estimates presented here it is difficult to draw any general conclusions about whether rates of price change are similar or dissimilar across different supermarket chains. This is because the index number formula chosen seems to be crucial in how this question is answered. Perhaps in practice, estimating price change over a longer period using the index number formula of interest may provide more useful information on rates of price change across supermarket chains rather than looking for regularities across different index number formulae. We now turn to the broader implications of this study.

8. Discussion

The use of scanner data to construct price indexes seems to be a mixed blessing. Price statisticians now have considerably more access to highly detailed data on consumer purchases than ever before, but the use of these data is not free from problems. One of the biggest issues is the increased volatility of price indexes relative to those constructed with more traditional data sources. This paper has made an attempt to find whether aggregation methods can be recommended through the use of hedonic regressions. Our results show that treating the same good as homogeneous across different stores which belong to the same chain, and in some

[16]An additional alternative approach was suggested by Silver (1999, 2009, 2010), which involves quality adjusting prior to the computation of the unit values. There are alternative ways of implementing this approach, which are left for future investigation.

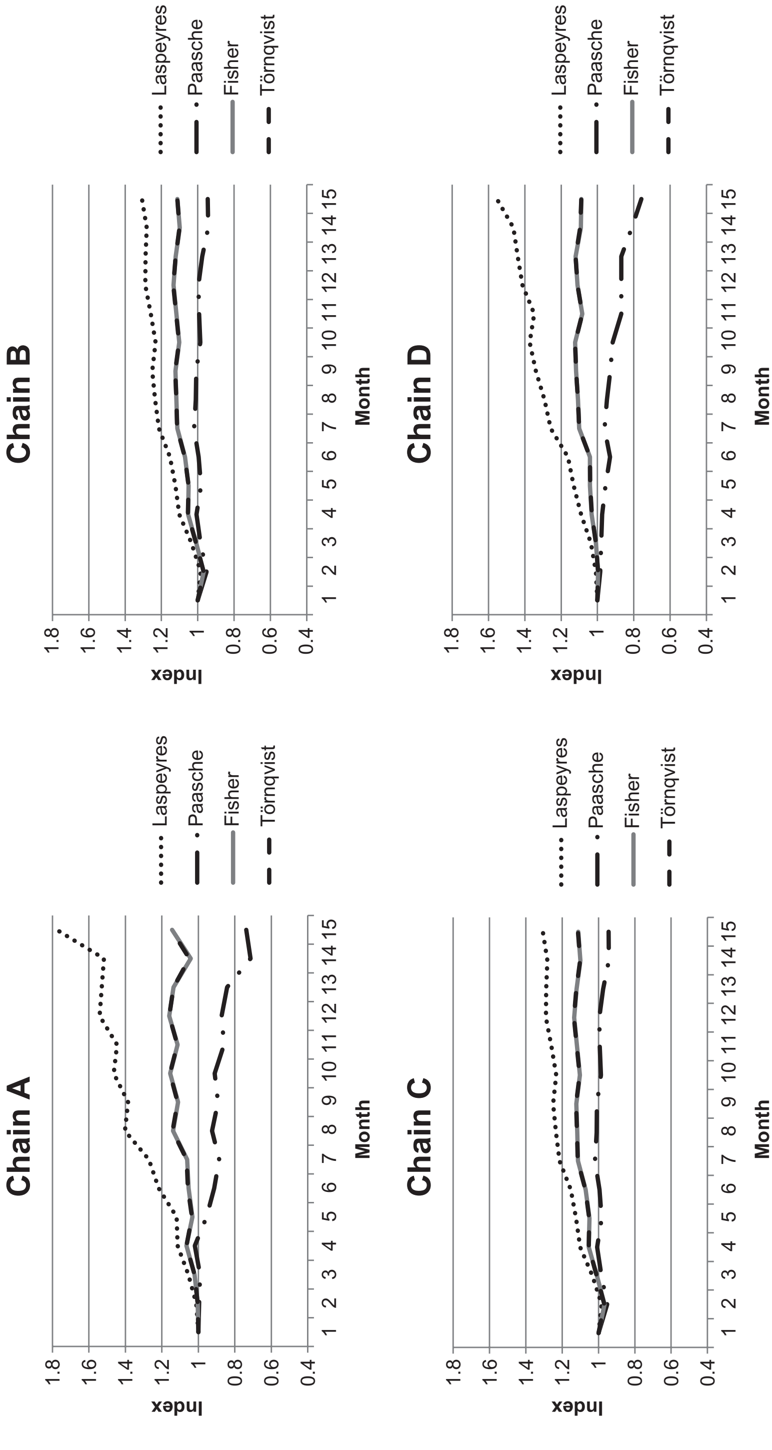

Figure 2. Monthly and Total Chained Estimates of Price Change for Each Supermarket Chain

circumstances across different chains, may be recommended. This is the first study of this type and, as such, it would seem premature to make generalizations about aggregation. Applying this analysis to a broader range of categories may help to determine whether any consistencies arise in the finding of homogeneity.

Statistical agencies could further use this type of analysis to inform decisions about sampling and how to set up sampling frames efficiently. For example, our analysis shows that choosing to sample from one store in Chain A or B may be enough to obtain representative prices for these supermarket chains. This does not appear to be the case for Chains C and D. Also, by identifying stores which do not have prices which are representative of general price levels within a chain, it can inform the understanding of price dispersion within chains.

Thus, this analysis framework may be used to better inform statistical agencies about both aggregation and sampling issues—two issues which are fundamental to the construction of price indexes. Importantly, it can also provide some insight into whether the implicit economic assumptions that are made when constructing estimates of price change are actually borne out when tested.

References

ABS (Australian Bureau of Statistics), *Australian Consumer Price Index: Concepts, Sources and Methods, Cat. No. 6461.0*, Australian Bureau of Statistics, Canberra, 2005.

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Caski (eds), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267–81, 1973.

Balk, B. M., "On the Use of Unit Value Indices as Consumer Price Subindices," Paper presented at the Fourth Meeting of the Ottawa Group at the International Conference on Price Indices, Washington, April 22–24, 1998.

Betancourt, R. and D. Gautschi. "The Economics of Retail Firms," *Managerial and Decision Economics*, 9, 133–44, 1988.

Bradley, R., "Pitfalls of Using Unit Values as a Price Measure or Price Index," *Journal of Economic and Social Measurement*, 20, 39–61, 2005.

Bradley, R., B. Cook, S. Leaver, and B. R. Moulton, "An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI," Paper presented at the Third Meeting of the Ottawa Group at the International Conference on Price Indices, Voorburg, April 16–18, 1997.

Dalen, J., "Computing Elementary Aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics*, 8, 129–47, 1992.

de Haan, J., "Estimating Quality-Adjusted Unit Value Indices: Evidence from Scanner Data," Paper presented at the SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver, June 30–July 3, 2004a.

———, "Hedonic Regression: The Time Dummy Index as a Special Case of the Imputation Törnqvist Index," Paper presented at the eighth Ottawa Group Meeting, Helsinki, August 23–25, 2004b.

———, Hedonic Price Indices: A Comparison of Imputation, Time Dummy and Other Approaches," Paper presented at the EMG Workshop, Sydney, Australia, December 12–14, 2007.

de Haan, J. and H. van der Grient, "Eliminating Chain Drift in Price Indexes Based on Scanner Data," *Journal of Econometrics*, 161, 36–46, 2011.

Diewert, W. E., "Exact and Superlative Indexes," *Journal of Econometrics*, 4, 115–45, 1976.

———, "Axiomatic and Economic Approaches to Elementary Price Indexes," NBER Working Paper No. W5104, 1995.

———, "Hedonic Regressions: A Review of Some Unresolved Issues," Presented at the Seventh Meeting of the Ottawa Group at the International Conference on Price Indices, Paris, May 27–29, 2003.

Feenstra, R. C. and M. D. Shapiro, *Scanner Data and Price Indexes*, University of Chicago Press, Chicago, IL, 2001.

Fox, K. J., "Non-Parametric Estimation of Technical Progress," *Journal of Productivity Analysis*, 10, 235–50, 1998.

Fox, K. J. and D. Melser, "Non-Linear Pricing and Price Indexes: Evidence and Implications from Scanner Data," *Review of Income and Wealth*, doi: 10.1111/roiw.12000, 2012.
Fuller, W. A., "Grafted Polynomials as Approximating Functions," *Australian Journal of Agricultural Economics*, 13, 35–46, 1969.
Goldberger, A. S., "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica*, 35, 464–72, 1968.
Halvorsen, R. and R. Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 70, 474–5, 1980.
Hawkes, W. J., "Reconciliation of Consumer Price Index Trends with Corresponding Trends in Average Prices for Quasi-homogeneous Goods using Scanner Data," Paper presented at the International Conference on Price Indices, Voorburg, April 16–18, 1997.
Hawkes, W. J. and F. W. Piotrowski, "Using Scanner Data to Improve the Quality of Measurement in the Consumer Price Index," in R. Feenstra and M. Shapiro (eds), *Scanner Data and Price Indexes*, University of Chicago Press, Chicago, IL, 2003.
Hill, R. J., "Superlative Index Numbers: Not All of Them Are Super," *Journal of Econometrics*, 130, 25–43, 2006.
ILO, *Consumer Price Manual: Theory and Practice*, International Labour Organization, Geneva, 2004.
Ivancic, L., W. E. Diewert, and K. J. Fox, "Scanner Data, Time Aggregation and the Construction of Price Indexes," *Journal of Econometrics*, 161, 24–35, 2011.
Jain, M. and R. Abello, "Construction of Price Indexes and Exploration of Biases using Scanner Data," Paper presented at the Sixth Meeting of the International Working Group on Price Indices, Canberra, April 2–6, 2001.
Kennedy, P. E., "Estimating with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 71, 801, 1981.
———, *A Guide to Econometrics*, 4th edition, MIT Press, Cambridge, MA, 1998.
Reinsdorf, M., "Changes in Comparative Price and Changes in Market Share: Evidence from the BLS Point-of-Purchase Survey," *Managerial and Decision Economics*, 13, 233–45, 1992.
———, "Price Dispersion, Seller Substitution and the U.S. CPI," Bureau of Labor Statistics Working Paper 252, March 1994.
———, "Using Scanner Data to Construct CPI Basic Component Indexes," *Journal of Business and Economic Statistics*, 17, 152–60, 1999.
Schwartz, G., "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–4, 1978.
Silver, M., "An Evaluation of the Use of Hedonic Regressions for Basic Components of Consumer Price Indices," *Review of Income and Wealth*, 45, 41–56, 1999.
———, "An Index Formula Problem: The Aggregation of Broadly Comparable Items," IMF Working Paper 09/19, 2009.
———, "The Wrong and Rights of Unit Value Indices," *Review of Income and Wealth*, 56, Special Issue 1, S206–23, 2010.
Silver, M. and S. Heravi, "The Measurement of Quality-Adjusted Price Changes," in R. Feenstra and M. Shapiro (eds), *Scanner Data and Price Indexes*, University of Chicago Press, Chicago, IL, 2003.
———, "The Difference between Hedonic Imputation Indexes and the Time Dummy Hedonic Indexes," *Journal of Business and Economic Statistics*, 25, 239–46, 2007.
Spulber, D. F., *Market Microstructure*, Cambridge University Press, Cambridge, 1999.
White, H., "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–38, 1980.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix A
Appendix B
Table B1: Model 1 Results: Base Chain = Chain A
Table B2: Model 2 Results: Base Chain = Chain A
Table B3: Index number estimates using different types of aggregation
Table B4: Monthly and total chained estimates of price change for each supermarket chain