# Constructing Divisia Indexes from Retail Scanner Data using Wavelet Methods

***Breige Allan-Greer***
***D.S. Prasada Rao***
***Christopher O'Donnell***
School of Economics
University of Queensland
St. Lucia, Brisbane, Australia

June 2004

The authors gratefully acknowledge the support of an Australian Research Council Linkage Grant. We are also greatly indebted to Bert Balk, Erwin Diewert, Kevin Fox and participants of the EMG Workshop in Sydney during 11-12 December, 2003 for their helpful comments.

# Constructing Divisia Indexes from Retail Scanner Data Using Wavelet Methods


## ABSTRACT

This paper represents an attempt to model continuous-time price and quantity functions from high-frequency scanner data with the aim of estimating Divisia indexes in a continuous-time framework. Wavelet-based interpolation methods are proposed and developed to estimate continuous-time price and quantity functions from which Divisia indexes can be calculated. The proposed methods are applied to weekly scanner data on potatoes and empirical Divisia price and quantity indexes are computed. Finally, the calculated Divisia price and quantity indexes are compared to common fixed-base and chained discrete-time approximations. The empirical Divisia indexes provide some support for the use of chained indexes.



**Journal of Economic Literature Classification Numbers: C43, C32**

**Keywords: Divisia indexes, scanner data, wavelets, price indexes.**

## 1. INTRODUCTION

The idea of representing aggregate value change as the product of a continuous-time price index and a continuous-time quantity index was first suggested and developed by French economist François Divisia in a series of papers around 1925[1]. These unique continuous-time indexes (which are often associated by name as Divisia indexes) have a long history particularly in the area of price and quantity measurement[2]. Generally, Divisia indexes are valued by statistical offices because they provide "an elegant logical justification" for the chaining of bilateral indexes over adjacent time periods[3]. Their real value though is much greater; under conventional neoclassical assumptions, Divisia indexes are the theoretical ideal in a continuous-time framework[4].

Divisia indexes are however difficult operationally because of the nature of the data underlying the indexes. Price and quantity data in their traditional forms are available only at regular discrete-time points like monthly intervals. As such, true continuous-time indexes cannot be calculated. Consequently, much of the research and literature relating to Divisia index numbers has been devoted to the derivation of discrete-time approximations[5]. Importantly, these discrete-time Divisia index approximations are viewed as "among the best available discrete-time statistical index numbers"[6].

The emergence of the relatively new data form of scanner data provides an exciting new opportunity to overcome the operational difficulties of Divisia indexes. Scanner data is becoming increasingly prevalent and many statistical offices are now integrating it with more traditionally collected data in order to derive consumer price indexes. This integration though is confined; scanner data is typically seen as an additional source of discrete data to improve monthly averages and weights[7]. The high frequency of observations in scanner data offers another opportunity – the possibility of modelling price and quantity observations as continuous functions of time thereby enabling the construction of continuous-time Divisia price and quantity indexes. This paper represents an attempt to model continuous-time price and quantity functions with the aim of estimating Divisia indexes.

---

[1] Divisia (1925).
[2] Balk (2000) and Diewert (1993) detail the historical development and context of Divisia indexes.
[3] Frisch (1936, 7). The use of Divisia indexes to rationalize chain indexes is further discussed in Balk (2000, 32).
[4] Hulten (1974) and Barnett and Serletis (2000, 113).
[5] For example, Törnqvist (1936), Theil (1967), Sato (1976), Vartia (1974, 1976) and Trivedi (1981).
[6] Barnett and Serletis (2000, 113).
[7] For example, Statistics Netherlands' use of scanner data outlined in Schut (2002).

The estimation of continuous-time functions from discrete-time data series is a vast statistical and mathematical field. In this paper, we propose the use of wavelet-based interpolation methods to construct continuous-time price and quantity functions from scanner data[8]. A wavelet is an oscillatory function with carefully constructed mathematical properties whose activity is limited to a 'small' interval of time. Wavelets act like building blocks and allow the complex scanner data to be decomposed and reconstructed at different positions and scales. The advantage that wavelet methods have over other methods (in particular Fourier methods) is that they can handle both stationary and non-stationary data series and are more equipped to deal with potentially volatile data[9]. Essentially, we are using wavelets to match a continuous-time model of price and quantity that is consistent with the high frequency observations of scanner data (and economic notions about how it should move). By choosing a wavelet appropriate to the price and quantity data and to the analysis goal of deriving Divisia index numbers, we show that a continuous-time price and quantity functions can be constructed from scanner data.

From these continuous-time functions, Divisia price and quantity indexes can be computed. We demonstrate these computations using retail scanner data on potatoes. We then examine how the calculated Divisia indexes compare to price and quantity indexes computed using more conventional statistical index number methods. This comparison is undertaken with both fixed-base and chained index numbers.

The paper is structured as follows. Section 2 examines Divisia index numbers and common discrete-time approximations. The continuous-time assumptions underpinning Divisia indexes are further explored in section 3. This section briefly addresses the question as to whether price and quantity data are intrinsically continuous in time. Section 4 then introduces the retail scanner data later analyzed in the paper's empirical work. An introduction to wavelets and a brief discussion of some specific wavelet methods follows in section 5. Section 6 then applies the wavelet theory to show how continuous-time price and quantity functions can be estimated from high-frequency scanner data. Empirical Divisia price and quantity indexes are subsequently calculated and then compared to common discrete-time approximations in section 7. Finally, conclusions and areas for further research are noted.

[8] Interpolation is the process of finding a function that passes through each of the observed points.
[9] Additional advantages are outlined in Appendix A.1.

## 2. DIVISIA PRICE AND QUANITY INDEXES

To begin, a basic overview of the Divisia index number approach is provided. Assume that a vector of unit prices and the corresponding vector of quantities for *N* commodities at time *t* can be considered respectively as $p^t \equiv ({p_1}^t, \mathrm{K}, {p_N}^t)$ and $q^t \equiv ({q_1}^t, \mathrm{K}, {q_N}^t)$. Commodity *n*'s $(n = 1, \mathrm{K}, N)$ value at time *t* is denoted by ${v_n}^t \equiv {p_n}^t {q_n}^t$ and the value of the aggregate at time *t* is expressed by $v^t \equiv p^t . q^t \equiv \sum_{n=1}^{N} {v_n}^t$. The value share of commodity *n* at time *t* is represented by ${s_n}^t \equiv \frac{{v_n}^t}{v^t} \equiv \frac{{p_n}^t {q_n}^t}{p^t q^t}$. Furthermore, given that the price and quantity of a particular commodity can continuously change, price and quantity are continuous functions of *t*, that is ${p_n}^t = p_n(t)$ and ${q_n}^t = q_n(t)$. It is also assumed that these price and quantity functions are differentiable[10].

The uniqueness of Divisia index numbers lies in their use of value shares to weight component growth rates to achieve an aggregate growth rate. Simply put, the value shares ${s_n}^\tau$ weight marginal products $\frac{d \ln {p_n}^\tau}{d\tau}$ or $\frac{d \ln {q_n}^\tau}{d\tau}$ for each commodity across all commodities. For example, the growth rate of a Divisia quantity index is $\sum_{n=1}^{N} {s_n}^\tau \frac{d \ln {q_n}^\tau}{d\tau}$.

As a continuous-time index, the *Divisia price index* is a line integral. It is therefore expressed as

$$\ln P_{Divisia}(t', t) \equiv \int_{t'}^{t} \sum_{n=1}^{N} {s_n}^\tau \frac{d \ln {p_n}^\tau}{d\tau} d\tau = \sum_{n=1}^{N} \int_{t'}^{t} {s_n}^\tau \frac{d \ln {p_n}^\tau}{d\tau} d\tau .$$

Similarly, the Divisia quantity index is

$$\ln Q_{Divisia}(t', t) \equiv \int_{t'}^{t} \sum_{n=1}^{N} {s_n}^\tau \frac{d \ln {q_n}^\tau}{d\tau} d\tau = \sum_{n=1}^{N} \int_{t'}^{t} {s_n}^\tau \frac{d \ln {q_n}^\tau}{d\tau} d\tau .$$

[10] A function is differentiable if it is continuous and smooth.

An alternative derivation of the Divisia line integral indexes is through analysis of the logarithm of the value ratio,

$$\ln \frac{v^t}{v^{t'}} \equiv \int_{t'}^{t} \frac{d \ln p^\tau . q^\tau}{d\tau} d\tau \equiv \int_{t'}^{t} \left( \frac{1}{p^\tau . q^\tau} \sum_{n=1}^{N} {q_n}^\tau \frac{d{p_n}^\tau}{d\tau} + \frac{1}{p^\tau . q^\tau} \sum_{n=1}^{N} {p_n}^\tau \frac{d{q_n}^\tau}{d\tau} \right) d\tau \equiv \ln P_{Divisia}(t,t') + \ln Q_{Divisia}(t,t').$$

Hence, Divisia indexes are the integral of value share weighted aggregates of the logarithmic rates of change in price and quantity respectively or alternatively, are the sum of each commodity's integral of value share weighted logarithmic price or quantity rate of change[11].

The Divisia index has a number of useful properties. Importantly, the Divisia approach decomposes the aggregate value change into two structurally similar parts. Accordingly, the price index is the exact dual of the quantity index and so satisfies Fisher's factor reversal test. This dual nature of prices and quantities is crucial for meaningful economic interpretation of index numbers. Divisia indexes are also consistent in aggregation which ensures that the behavioural properties suggested by the disaggregate data are likewise represented at the aggregate level. In fact, Divisia indexes satisfy most requirements of a bilateral index including transitivity and monotonicity[12]. Consequently, Divisia indexes are significant even under standard index number criteria.

It can even be argued that a Divisia index is "the approach" in continuous-time. Hulten (1973) proved that the continuous-time Divisia index is an exact index for any consistent aggregator functions under conventional neoclassical assumptions. These assumptions are that the aggregator function is blockwise homothetically weakly separable[13]. As such, "no index number can be better than the Divisia in continuous-time"[14]. The work of Malaney and Weinstein outlined in Balk (2000) further substantiates this. Using the sole assumption of rational economic behaviour, they provided a dynamic (in the sense of preferences) economic interpretation of Divisia indexes. Such work furthers the assertion that Divisia indexes are theoretically ideal index numbers.

[11] Diewert (2001, 80-82) outlines another alternative derivation using the economic approach.
[12] Balk (2000, 11-16).
[13] Barnett and Serletis (2000, 39). Static preferencing is the main obstacle in meaningful economic interpretation of Divisia indexes.
[14] Barnett and Serletis (2000, 39).

There are practical problems in the implementation of the Divisia indexes – the continuous-time functions $p^{\tau}$ and $q^{\tau}$ are not known. This problem has led to discrete-time approximations of Divisia indexes. The most popular discrete-time approximation is the Törnqvist-Theil index defined in Table 3.1[15]. It can be derived using a variety of methods; simplest of which are numerical integral approximations like Simpson's rule. Under certain assumptions[16], the Sato-Vartia index[17] expressed in Table 3.1 is also a discrete-time approximation of the Divisia index. Fixed base Laspeyres and the Paasche index numbers[18] denoted in Table 3.1 can also approximate Divisia indexes using forward and backward differences.

Table 3.1. Discrete-time approximations of Divisia price and quantity indexes

| Discrete-time approximation | *Price index* | *Quantity index* |
|---|---|---|
| *Törnqvist-Theil* | $P_{TT}(t,0)=\prod_{n=1}^{N}\left(\frac{{p_n}^t}{{p_n}^0}\right)^{\frac{{s_n}^t+{s_n}^0}{2}}$ | $Q_{TT}(t,0)=\prod_{n=1}^{N}\left(\frac{{q_n}^t}{{q_n}^0}\right)^{\frac{{s_n}^t+{s_n}^0}{2}}$ |
| *Sato-Vartia*[19] | $\ln P_{SV}(t,0)=\frac{\sum_{n=1}^{N}L({s_n}^t,{s_n}^0)\ln({p_n}^t/{p_n}^0)}{\sum_{n=1}^{N}L({s_n}^t,{s_n}^0)}$ | $\ln Q_{SV}(t,t')=\frac{\sum_{n=1}^{N}L({s_n}^t,{s_n}^0)\ln({q_n}^t/{q_n}^0)}{\sum_{n=1}^{N}L({s_n}^t,{s_n}^0)}$ |
| *Fixed-base Laspeyres* | $P_{Laspeyres}(t,0)=\frac{\sum_{n=1}^{N}{p_n}^t.{q_n}^0}{\sum_{n=1}^{N}{p_n}^0.{q_n}^0}$ | $Q_{Laspeyres}(t,0)=\frac{\sum_{n=1}^{N}{p_n}^0.{q_n}^t}{\sum_{n=1}^{N}{p_n}^0.{q_n}^0}$ |
| *Fixed-base Paasche* | $P_{Paasche}(t,0)=\frac{\sum_{n=1}^{N}{p_n}^t.{q_n}^t}{\sum_{n=1}^{N}{p_n}^0.{q_n}^t}$ | $Q_{Paasche}(t,0)=\frac{\sum_{n=1}^{N}{p_n}^t.{q_n}^t}{\sum_{n=1}^{N}{p_n}^t.{q_n}^0}$ |

[15] Törnqvist (1936) and Theil (1967). Its popularity lies in its classification as a superlative index number: refer to Diewert (1976).
[16] Assumes a rational economic agent with time-invariant CES preferences (Balk, 2000).
[17] Sato (1976), Vartia (1974, 1976).
[18] Refer to Diewert (2001, 8-10).
[19] $L$ is the logarithmic mean, $L(a,b)=\begin{cases} a & a=b \\ (a-b)/\ln(a/b) & a\neq b \end{cases}$.

Moreover, the transitive character of Divisia indexes means they can be approximated by chained indexes. Any bilateral index can be chained to approximate the corresponding Divisia index[20]. The smaller the time interval spanned in the chaining the better the approximations[21]. Common bilateral index numbers used in chaining are the Törnqvist-Theil and Fisher index numbers.

All of the above discrete-time indexes lack properties integral to the 'exact' proof of Hulten (1973). In other words, it cannot be assumed that the aggregator function is blockwise homothetically weakly separable. Therefore, the discrete-time indexes are incapable of accurately representing the value of every economic aggregate. There will always be a degree of error. Though the magnitude of error may not always be significant, it is restrictive in an economic sense. The error originates from the character of Divisia indexes as line integrals which means that they are dependent on the path over which integration is taken. As the frequency of observations increases, however, the impact of path dependency decreases and further decreases with careful estimation of continuous-time price and quantity functions.

## 3. THE CONTINUOUS-TIME FRAMEWORK

It is clear from the preceding section that the assumption price and quantity are continuous functions of time is integral to the theory of Divisia indexes. It has though obscured Divisia index number theory. This is because the continuous-time assumption is perceived to lack practicality[22]. The purpose of this section is to briefly consider the continuous-time assumption and outline some arguments to support its use in practice.

The framework from which most index number theories derive economic meaning is a continuous one. Two fundamental assumptions of economic theory are that price and quantity interact continuously and that consumers exhibit optimizing behaviour. So as prices change, consumers instantaneously adjust the quantities demanded continuously so as to continually engage in optimizing behaviour. Economic interpretation of index numbers is reliant on this framework of continuous preferences and continual optimization. Accordingly, the continuous framework of Divisia indexes is not novel.

[20] Refer to Frisch (1936, 8), Balk (2000) and Diewert (2001).
[21] This is true to a point. At very short intervals, it is likely additional bias may be introduced.
[22] For example, Diewert (2001, 17).

What is unique about Divisia indexes is the explicit assumption that time is a continuous variable, or rather that price and quantity are continuous functions of *time*. Economic theory makes no explicit reference to time. Many standard index numbers are based on discrete-time because it is argued that economic data are collected at discrete-time intervals in practice. The question arises then whether it is practical for prices to be modelled as continuous functions of time[23].

We contend that it is natural and reasonable to treat price as a continuous-time phenomenon[24]. That is not to say that prices must change value all the time, or that they cannot change value at predetermined discrete instants. We simply recognize that in practice prices react continuously to the events and consumer behaviour of the real world. Continuous-time allows for the interaction of price with other economic variables to be fully accounted for.

Moreover, prices are not just the result of a few individual economic decisions. Rather, they are the result of multitudes of economic decisions by a multitude of economic agents at a multitude of different time points. In the context of such aggregation, it only seems natural to treat price as continuous in time. Only by regarding price in this manner, can the movements of price really be traced.

The deficiencies of a discrete-time price function also lend weight to a continuous-time characterization of price. A discrete-time price function implies that prices can only change value at regularly spaced and exact time intervals. Prices though can change value at irregular and shorter time intervals than current regular discrete measurements capture (as scanner data demonstrates). Furthermore, the implication that price is a discrete-time function merely because we are only able to take discrete-time measurements is also flawed. Firstly, some forms of scanner data provide continuous-time measurements. Secondly, regardless of what we are capable of observing or recording, the real world is continuous in time. While acknowledging this may pose challenges, any theoretical or practical characterization of price should attempt to model what actually occurs, which is something that a continuous-time function is better able to do.

[23] If price is a continuous function of time, then quantity must also be a continuous-time function by way of the assumption that price and quantity interact continuously.
[24] Bergstrom (1996, 3-6) and Gandolfo (1997, 547-579) detail a number of specific advantages of continuous-time. This paper only offers more general advantages.

Sometimes observed price data has discontinuities and these discontinuities are used to dismiss continuous-time price characterizations. For example, the closing price on one trading day does not match the starting price on the next trading day. This jump in the observed data does not however represent a discontinuity in the price function. The price function underlies the transaction price data. The real price has reacted to an external economic variable or variables and changed value during an unobserved period. While it is possible that the movement in price is over a very short interval, the price function is nonetheless continuous in time.

The rationale for index numbers also provides support for the continuous-time price and quantity assumptions. A price index is a system by which changes in the price of some group of commodities and the rate at which they change value can be compared or measured. A quantity index is similarly, a system by which changes in the quantity demanded of some group of commodities and the rate at which they change value can be compared or measured. The rationale for index numbers is essentially to model the movements of price and quantity. Only continuous-time functions fully account for this movement. Discrete-time functions fail to account for the movement between the discrete-time points which can be quite substantial in some instances.

The continuous-time price and quantity assumptions made in deriving Divisia indexes, we therefore believe, have sufficient theoretical and practical grounding. Moreover when analysing high frequency of observations in scanner data, we argue that the continuous-time assumptions are particularly appropriate.

## 4. RETAIL SCANNER DATA

Scanner data or high frequency data, as it is sometimes referred to, typically has two forms – at its most basic, it contains the electronic records of every individual transaction including the time, product code, transaction price and associated quantity sold ('scanner data'); or in its more general form, it contains the total dollar sales and total quantity sold for each particular product code over particular aggregate time periods such as by weeks, months or quarters ('high-frequency data')[25].

[25] The labels associated with each form for the purposes of this paper are enclosed in the brackets. Feenstra & Shapiro (2003) similarly used the term high-frequency to refer to weekly scanner data.

The retail data set studied in this paper is high-frequency data. It is drawn from the transaction records of a large independent fresh produce retailer in Australia. Each cross-section provides the weekly total quantity sold per kilogram and weekly total sales on a dollar basis for every product at each outlet. The dollar-based average price per kilogram per week used in the indexes has been calculated for each outlet as a quantity weighted average price from the weekly sales and quantity figures. The weekly data covers an eight year period beginning in July 1993 (a total of 418 weeks). The relevant subset of data used to calculate Divisia indexes is that of three potato varieties at three outlets in Sydney. Tables 4.1 and 4.2 below provide some insight into this subset of data.

Table 4.1. Summary statistics of the (average) weekly price ($/kg) across the three outlets

| | Variety 1 | | | Variety 2 | | | Variety 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Outlet | A | B | C | A | B | C | A | B | C |
| Min | 0.4075 | 0.4075 | 0.5155 | 0.0900 | 0.3200 | 0.2600 | 0.4400 | 0.4300 | 0.5200 |
| 1st Quartile | 1.2900 | 1.2900 | 1.2900 | 0.7900 | 0.8900 | 0.8725 | 1.2900 | 1.2900 | 1.2900 |
| Mean | 1.5491 | 1.5491 | 1.5781 | 0.9934 | 1.1386 | 1.0174 | 1.5534 | 1.5926 | 1.5799 |
| Median | 1.4901 | 1.4901 | 1.5954 | 0.9900 | 0.9900 | 0.9900 | 1.5950 | 1.6900 | 1.6900 |
| 3rd Quartile | 1.7900 | 1.7900 | 1.9752 | 1.2900 | 1.4900 | 1.2500 | 1.9900 | 1.9900 | 1.9900 |
| Max | 2.4901 | 2.4901 | 2.4902 | 1.6900 | 1.9900 | 1.6900 | 2.4900 | 2.8600 | 2.4900 |
| St Dev | 0.3622 | 0.3622 | 0.3974 | 0.3334 | 0.3933 | 0.2983 | 0.3833 | 0.4173 | 0.3888 |

Table 4.2. Summary statistics of weekly quantity sold (kg) across the three outlets.

| | Variety 1 | | | Variety 2 | | | Variety 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Outlet | A | B | C | A | B | C | A | B | C |
| Min | 51.50 | 74.80 | 58.70 | 245.60 | 390.40 | 355.20 | 99.50 | 307.90 | 125.90 |
| 1st Qu. | 147.38 | 365.90 | 223.63 | 595.73 | 673.05 | 753.85 | 231.53 | 607.53 | 378.95 |
| Mean | 228.53 | 524.81 | 366.10 | 945.38 | 1122.68 | 1118.79 | 340.67 | 807.33 | 516.64 |
| Median | 182.80 | 451.90 | 306.30 | 772.15 | 874.20 | 949.95 | 282.85 | 702.60 | 451.10 |
| 3rd Qu. | 249.63 | 596.80 | 416.93 | 952.15 | 1065.48 | 1347.23 | 341.88 | 825.63 | 532.98 |
| Max | 2236.00 | 2971.20 | 1719.80 | 5338.90 | 4418.70 | 4002.50 | 2371.20 | 3123.60 | 1930.40 |
| St Dev | 213.16 | 327.10 | 250.47 | 646.61 | 801.54 | 534.91 | 257.96 | 436.71 | 285.99 |

This particular subset of data has been chosen for its relative simplicity. Potatoes are a stable part of the diet of many Australian consumers and as such there is no need to estimate missing observations in the data subset. Potatoes are also grown almost year round in Australia and consequently, issues relating to seasonality are minimized. Furthermore, the potato varieties are closely related to the extent that they can be considered close substitutes.

Scanner data and even high frequency data sets like the one studied here cover the universe of transactions made compared to only samples typically used in calculating indexes. The additional information gained in scanner data offers great benefits. The high frequency of scanner data provides an exciting opportunity to estimate continuous-time price and quantity functions and hence enable the computation of empirical continuous-time Divisia price and quantity indexes.

## 5. REPRESENTATION OF CONTINUOUS-TIME PROCESSES BY WAVELETS

### 5.1. Introduction to wavelet analysis[26]

A *wavelet* is an oscillatory function whose energy is concentrated in time. Simply, wavelets are small waves or localized wiggles. Figure 5.1 compares an example of a 'big' wave, the sine wave, having infinite energy and equal amplitude with an example of a wavelet, the Morlet wavelet, which has a limited duration with an average value of zero.

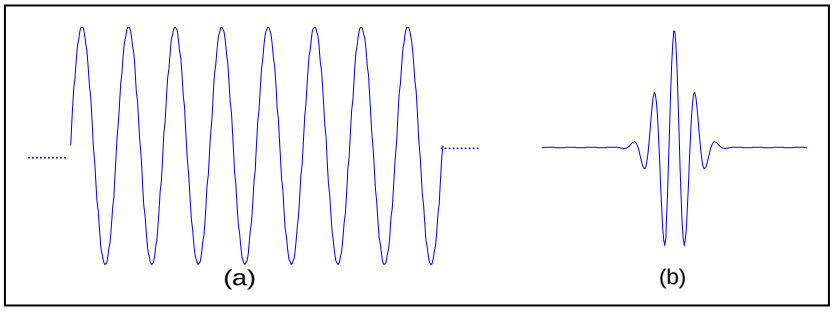



Figure 5.1: (a) sine wave and (b) Morlet wavelet

[26] Readers seeking a more detailed introduction to wavelets or further detail on any of the wavelet methods discussed are referred to Ogden (1997), Vidakovic (1999) and Percival & Walden (2000) (statistical); or Gencay, Selcuk & Whitcher (2002) (economic).

Mathematically, a wavelet is any real-valued function[27] $\psi(t)$ with the following two properties:

1. $\psi(t)$ integrates to zero: $\int_{-\infty}^{\infty} \psi(t)dt = 0$; and
2. $\psi(t)$ is square integrable: $\int_{-\infty}^{\infty} \psi^2(t)dt = 1$[28].

Many wavelet functions have already been developed which are suitable for present purposes therefore it is unnecessary to consider the construction of any wavelets in this paper.

A wavelet function (also known as a *mother wavelet*) can be dilated and translated to construct a family of wavelets. Dilation ($j$) represents the degree of compression or scale of the wavelet while translation ($k$) determines the time location of the wavelet. As parameters $j$ and $k$ can vary continuously or discretely, the construction of wavelet families can be defined in a continuous and discrete framework as follows,

- *continuous construction* $\psi_{j,k}(t) = j^{-1/2}\psi(\frac{t-k}{j})$ where $j \in \boldsymbol{R}^+$ and $k \in \boldsymbol{R}$
- *discrete construction* $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ where $j \in \boldsymbol{Z}^+$ and $k \in \boldsymbol{Z}$.

Figure 5.2 demonstrates the effect of changing $k$ and $j$ on the *Haar mother wavelet*. The Haar mother wavelet is the simplest wavelet and is defined as $\psi(t) \equiv \begin{cases} 1 & 0 < t \le 0.5 \\ -1 & 0.5 < t < 1 \\ 0 & otherwise \end{cases}$. The mother wavelet is increasingly compressed as $j$ increases and is increasingly stretched as $j$ decreases. A large $j$ corresponds mainly to higher frequencies while a small $j$ corresponds to lower frequencies[29]. The effect of changing $k$ is to slide the wavelet along the time axis[30].

[27] A wavelet function can be real or complex-valued but in this paper only real-valued wavelet functions shall be considered. Refer to Percival & Walden (2000) for more information.
[28] Percival & Walden (2000, 3). Depending on the wavelet chosen, second order moments may not be possible or alternatively, higher order moments may also be required to equate to zero. These mathematical properties relate to continuous wavelets. Analogous properties for discrete wavelets (sometimes known as discrete wavelet filters) are summation to zero and unit energy.
[29] A high frequency means the variable is changing value rapidly while a low frequency refers to a slow change in value. High frequencies/low scales relate to a detailed view of a data series (that usually lasts a relatively short time) whereas low frequencies/high scales correspond to a more global view of a data series.
[30] Ogden (1997, 8).

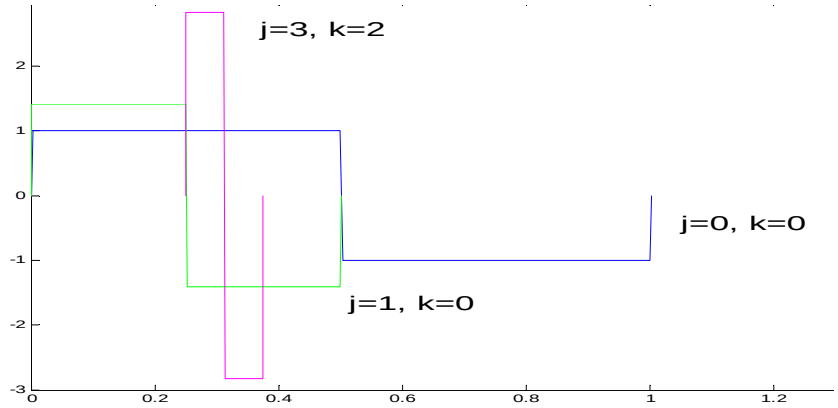



Figure 5.2: Dilations and translations of a Haar mother wavelet.

The resulting set of translated and dilated wavelets can then be used as building blocks to describe a data series provided that $\psi(t)$ satisfies the admissibility condition (in the continuous case) or is orthogonal to even shifts (in the discrete case). The admissibility condition is satisfied if $0 < C_\psi < \infty$ where $C_\psi = \int_0^\infty |\Psi(w)|^2 / w\, dw$ and $\Psi(w)$ is the Fourier transform of $\psi(t)$[31]. $C_\psi$ is used to normalize a continuously constructed representation. Turning to the discrete case, a $\psi(t)$ is orthogonal to its even shifts if $\sum_{l=0}^{L-1} \psi(l)\psi(l+2n) = 0$ for all non-zero integers $n$[32]. Orthogonality in turn establishes orthonormality which enables a discrete type representation to be expressed as a linear combination of a set of wavelets. All mother wavelets used in this paper satisfy the relevant conditions.

Using the family of wavelets, it is possible to represent a data series $f(t)$ in following two ways:

***1. Wavelet decomposition***

- *continuous construction* $$d_{j,k}^{CWT} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt = j^{-1/2} \int_{-\infty}^{\infty} f(t)\psi(\frac{t-k}{j})dt \qquad (1)$$
- *discrete construction* $$d_{j,k}^{DWT} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt = 2^{j/2} \int_{-\infty}^{\infty} f(t)\psi(2^j t - k)dt \qquad (2)$$

[31] $\Psi(w) \equiv \int_{-\infty}^{\infty} \psi(u) e^{-i2\pi wu} du$ .

[32] Percival & Walden (2000, 69).

These constructions are known as the *continuous wavelet transform* (CWT) and the *discrete wavelet transform* (DWT) respectively. The result of applying the CWT to $f(t)$ is the CWT coefficients $d_{j,k}^{CWT}$ and similarly, the DWT result is the DWT coefficients $d_{j,k}^{DWT}$. Wavelet transforms can be thought of as measures of the similarity between translated and scaled versions of the mother wavelet and the data series under analysis. They are akin to Fourier transforms in that they are the projection of $f(t)$ onto sets of functions except that here wavelet functions $\psi_{j,k}(t)$ are used instead of the functions $e^{\text{int}}$.

It is important to note that the use of continuous and discrete with respect to wavelet transforms does not relate to the nature of the data. Rather, what is continuous about the CWT is the continuous set of scales at which it operates and its continuous shifting across the data series. The discreteness of the DWT relates to the dyadic scales and positions.

In fact, the DWT can be thought of as the discrete sampling of CWT at dyadic scales and positions[33]. By looking at dyadic scales, the DWT assumes that the number of observations $N$ can be expressed in dyadic terms, that is, as a power of 2 so $N=2^J$ for some $J$. This assumption is based on the argument that CWT coefficients are likely to contain redundant data as the CWT looks at time and frequency continuously (that is, creates a two-dimensional representation of the data series). The DWT seeks to minimize this redundancy by dyadic sampling of the CWT. While sampling at just the dyadic scales and positions might seem a drastic reduction, it is still at a critical sampling level[34]. This means that there is no loss of information in going from the CWT to the DWT and we are still able to derive a unique function representation[35]. The DWT is just a succinct characterization of the CWT.

[33] The DWT can also be derived independently of the CWT. Refer to Percival & Walden (2000); Vidakovic (1999).
[34] Critical sampling is the minimum sampling that ensures no loss of information and that the function can be uniquely decomposed and reconstructed. Refer to Vidakovic (1999,50).
[35] This is true for most data series including the scanner data studied in this paper. In the event that it cannot discrete wavelet packet transforms can be used. Refer to Vidakovic (1999) and Percival & Walden (2000) for more information.

***2. Wavelet reconstruction[36]***

- *continuous construction* $$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{0}^{\infty} d_{j,k}^{CWT} j^{-1/2} \psi\left(\frac{t-k}{j}\right) \frac{dj}{j^2} dk \quad (3)$$

- *discrete construction* $$f(t) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}^{DWT} 2^{j/2} \psi(2^j t - k) \quad (4)$$

These representations are the projection of the relevant wavelet transform coefficients $d_{j,k}$ onto a wavelet family $\psi_{j,k}(t)$ over the range of values of *j* and *k*. Importantly, all the information in the original data series is preserved in both these representations. Given we shall be analysing finite scanner data later in the paper; the discrete construction (4) applies to finite data and hence uses finite parameters[37]. A more formal definition would extend the range of *j* and *k* to infinity.

These two methods of wavelet decompositions and reconstructions are simply two different "representations for same mathematical entity"[38]. For the purposes of estimating functions using high-frequency scanner data, the discrete wavelet reconstruction representation (4) is the most appropriate representation.

This concludes the brief introduction to wavelet analysis. Actual implementation of wavelet methods through computational algorithms is slightly more complicated and further information on these matters is briefly provided in Appendices A.2 and A.3.

## 5.2 Specific wavelet methods

A number of practical issues are raised by the brief examination of wavelets, specifically that,

1. *J* may not be a dyadic power as required by the DWT (Partial DWT and Padding); or
2. How to remove noise from observed data (Thresholding); or
3. Estimation of a derivative (Vaguelette-Wavelet Decomposition).

[36] These reconstructions are more commonly referred to as inverse continuous wavelet transforms and inverse discrete wavelet transforms.
[37] This was done to limit confusion later in the paper.
[38] Percival & Walden (2000, 15).

The purpose of this section is to examine the specific wavelet methods that overcome these potential problems in light of the analysis goal and data. All calculations for the remainder of the paper are performed using Matlab and its associated wavelet toolbox.

***Partial DWT***

Recall that the DWT is based on the premise that the sample size is a power of two. This assumption is very restrictive and a number of modified DWTs have been developed to overcome it[39]. The simplest of these is the partial DWT. It instead assumes that $N$ is an integer multiple of $2^{J_p}$ where $J_p < J$. The representation of the function analogous to (4) using the partial DWT is

- *discrete construction* $$f(t) = \sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} d_{j,k}^{DWT} 2^{j/2} \psi(2^j t - k) \qquad (5)$$

Like the DWT, the partial DWT still preserves all the information about the data series.

Another popular modified DWT is the maximal overlap DWT (MODWT) which only samples the CWT at dyadic scales. It can handle any sample size but it is highly redundant and consequently the partial DWT in conjunction with padding is the preferred option.

***Padding***

If the sample size cannot be expressed in dyadic terms or as an integer multiple of a dyadic term, then the observed series must be 'padded' at the boundaries to increase the number of observations to the nearest dyadic term or dyadic multiple[40]. The extra observations are added on both sides and are removed from the series after wavelet analysis. By increasing the same size, the DWT or partial DWT may now be able to be applied. Given the cyclical nature of the data, minimal periodic padding is advocated in the present paper. There are alternative padding methods like symmetric padding and zero padding but these options are not preferred for the fresh produce scanner data. Checks were performed to ensure that there was no difference in the results for the same series truncated to a dyadic length and a non-dyadic length for either price or quantity data.

Hence, if sample size $N = 2^J$ for some $J \Rightarrow$ use the DWT reconstruction (4)*;*
$= b.2^J$ for some $J$ and $b \in \boldsymbol{Z}^+ \Rightarrow$ use the partial DWT reconstruction (5); or
$= a+b.2^J$ for some $J$ and $a,\ b \in \boldsymbol{Z}^+ \Rightarrow$ use padding to increase $N$ to the nearest $b.2^J$ and then use the partial DWT reconstruction (5).

[39] There are many modified DWTs refer to Ogden (1997), Vidakovic (1999) and Percival & Walden (2000).
[40] Alternative methods are discussed in Percival & Walden (2000, 141-145).

***Thresholding***

Every observed data series is considered to contain observations from the true function plus a noise series. The aim is to recover the underlying function from the noisy data. Thresholding is a technique developed by Donoho and Johnstone (1994, 1995, 1998) to 'denoise' the wavelet transform coefficients leaving only those of the true function which can in turn be reconstructed into series[41]. This is achieved without assuming any particular parametric structure for the underlying function though it is assumed that the noisy observations are independent and identically distributed normal random variables with zero mean and variance $\sigma^2$. It operates from the basis that only a few 'large' wavelet coefficients contain information about the function while all other 'small' wavelet coefficients contribute noise. Hence, the goal of thresholding is to determine and retain the large coefficients and zero out as many possible small coefficients. By doing so, we obtain an approximate wavelet representation of the true underlying function.

Thresholding also has an important consequence in light of our analysis goal. It has the effect of smoothing out the data which is important given the differentiable assumption required to construct Divisia indexes.

Thresholding involves two steps – determining what is 'large' (setting a threshold level $\lambda$) and then determining the scheme to be applied to the wavelet coefficients that zeroes out the 'small' coefficients (applying a threshold scheme).

<u>Applying a threshold scheme</u>

Two popular rules developed by Donoho and Johnstone (1994) involving the killing or shrinking of smaller coefficients and the keeping of larger coefficients. These schemes are defined as follows,

1. Hard thresholding - 'keep or kill' rule - $\delta_\lambda^H(d_{j,k}) = \begin{cases} d_{j,k} & \text{if } |d_{j,k}| > \lambda \\ 0 & \text{otherwise} \end{cases}$

2. Soft thresholding - 'shrink or kill' rule

$$\delta_\lambda^S(d_{j,k}) = sign(d_{j,k})(|d_{j,k}| - \lambda)_+ \quad \text{where} \quad sign(d_{j,k}) = \begin{cases} +1 & \text{if } d_{j,k} > 0 \\ 0 & \text{if } d_{j,k} = 0 \\ -1 & \text{if } d_{j,k} < 0 \end{cases} \quad \text{and} \quad (x)_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

[41] Thresholding builds on a branch of modern statistical theory known as shrinkage estimation. Donoho and Johnstone developed the technique from a rigorous statistical point of view.

Hard thresholding sets to zero elements whose absolute values are lower than a predetermined threshold level. Soft thresholding likewise sets to zero elements whose absolute values are lower than a predetermined threshold level but also shrinks the non-zero coefficients towards 0. The soft thresholding approach is adopted in paper as hard-thresholding creates a discontinuity at $d_{j,k} = \pm\lambda$.

While it is noted that there are move advanced threshold rules involving two threshold parameters can be adopted, the soft-thresholding rule is adequate for present purposes[42].

Setting a threshold level λ

The proper choice of threshold level λ involves careful consideration. Choosing a very large threshold will make it very difficult for a coefficient to be judged 'large' and included in the reconstruction, consequently resulting in over-smoothing. Conversely, a very small threshold value will allow numerous coefficients to be included in the reconstruction resulting in an under-smoothed estimate. Three popular threshold levels are:

1. Universal threshold (Donoho and Johnstone, 1994)

$\lambda_U = \sigma\sqrt{2\ln N}$ where σ is the standard deviation of the noise data

The rationale is to remove all wavelet coefficients smaller than the expected maximum of an assumed independent and identically distributed normal noise sequence of the same size. In practice, σ is usually unknown and hence it is typically estimated by the median absolute deviation (MAD) of the estimated wavelet coefficients from the median at the finest level divided by 0.6475[43]. A very simple alternative is to assume that σ = 1, that is, Gaussian noise. The universal threshold is typically very efficient at removing noise, however, in its eagerness it also removes part of the underlying function. As a consequence, it may tend to over-smooth the data. Thus, the universal threshold is not considered to be the best choice given our analysis goal.

[42] For example, firm thresholding.

[43] MAD=$\frac{median(|d_{J-1} - median(d_{J-1})|)}{0.6745}$ : Ogden (1997, 132).

2. Minimax threshold (Donoho and Johnstone, 1994)

$$\lambda_{MIN} = 0.3936 + 0.1829(\log_2 N)$$

The rationale for the minimax threshold is to yield the yield minimax performance for mean square error against an ideal procedure[44]. While the minimax threshold is like the universal threshold in that it is predetermined, it is not as zealous as the universal threshold in its removal of noise. It does not then oversmooth the data and retains more of the underlying function. The minimax threshold is therefore a reasonable possibility for a threshold leel.

3. A threshold based on Stein's unbiased risk estimator (SURE) (Donoho and Johnstone, 1995)

$$\lambda_{SURE,j} = \arg\min_{\lambda \geq 0}(N - 2 \cdot \#\left\{d_{j,k} : \left|d_{j,k}\right| \leq \lambda\right\} + \sum(\min(\left|d_{j,k}\right|, \lambda)^2)$$

The rationale for the SURE threshold is to minimize the unbiased estimator of risk developed by Stein (1981)[45]. Unlike the other thresholds presented in this paper, the SURE threshold is data and level dependent. As a result, the SURE threshold is the least efficient method (of those presented) for the removal of noise. It is however not expected that the scanner data will contain excessive amounts of noise; so this threshold is the preferred choice for later analysis.

The choice of threshold level can have a significant impact on the reconstructed function. For this reason, calculations of Divisia indexes will entail all four threshold levels. The SURE threshold is preferred to the minimax threshold which is preferred to the universal threshold. This order of preferences is reflected in practice as shown in Figure 5.3. The original observations are shown in blue dots and the reconstructed threshold observations are shown in red. With the progression from universal to SURE thresholding results, less noise is removed (less blue can be seen).

[44] Gencay, Selcuk & Whitcher (2002, 213-214).
[45] Ogden (1997, 146) illustrates a clear relationship between the SURE threshold and Akaike's Information Criterion.

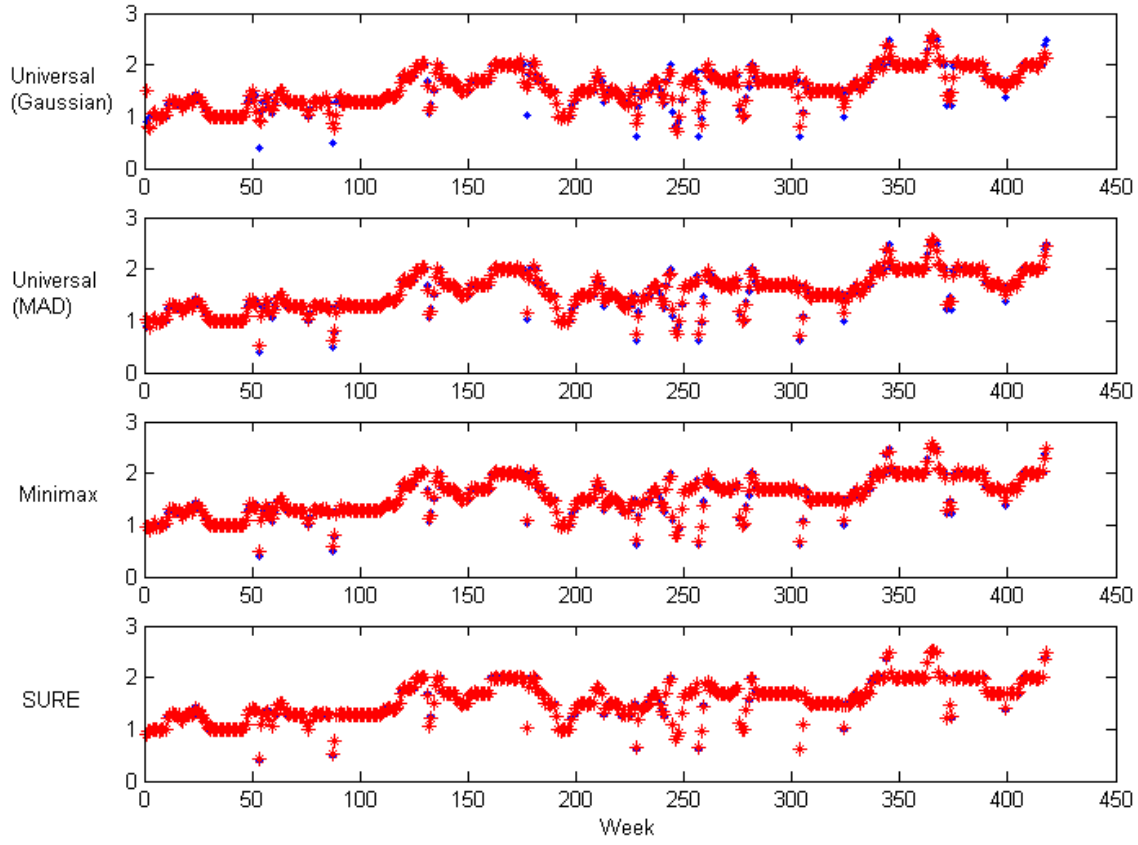



Figure 5.3: Comparison of threshold levels using soft thresholding rule for a sample price series. The blue dots are a sample high-frequency price series. The red asterisks are the high-frequency price series after thresholding.

### *Vaguelette Wavelet Decomposition*[46]

Interestingly, the set of discrete type wavelets constructed by dilating and translating a mother wavelet form a basis for a function space[47]. This wavelet basis can be used instead of an eigenfunction basis to calculate the solutions to linear inverse problems. In particular, Abramovich and Silverman (1998) and Donoho (1992) suggest that 'waveletized' methods akin to the standard singular value decomposition can be used for solving the derivative of an observed data series.

[46] Vaguelettes are similar to wavelets; they are represented as dilations and translations of a single mother vaguelette however the set of vaguelettes are only near orthonormal. Often described as 'almost wavelets'.
[47] A collection of functions forms a basis for a function space if the functions span the entire space and are linearly independent.

Using vaguelette–wavelet decomposition (VWD), Abramovich and Silverman (1998) define the derivative of an observed data series with respect to time as,

$$\frac{df(\tau)}{d\tau} = \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \delta_\lambda^S(d_{j,k}^n)\psi'_{j,k}(\tau) = \sum_{j=0}^{J-1} 2^{\frac{3}{2}j} \sum_{k=0}^{2^j-1} \delta_\lambda^S(d_{j,k}^n)\psi'_{j,k}(2^j\tau - k)\,.$$

The explicit calculation of $\psi'_{j,k}(\tau)$ is not easily achieved. Compactly supported wavelets are suggested by Abramovich and Silverman (1998) and Donoho (1992) in their 'waveletized' decompositions to reduce the computational strain[48]. Kolaczyk (1994) suggests converting the data to the frequency domain using the fast Fourier transform (FFT) and applying the derivative operator by scaling the FFT coefficients, then computing the inverse FFT. This is still a demanding task.

Abramovich and Silverman (1998) instead offer as an alternative the following approach:

1. calculate the DWT
2. threshold the wavelet coefficients
3. reconstruct the series using the inverse DWT (4)
4. apply the derivative operator to the reconstructed result of step 3 which can be done "by stable numerical methods"[49]

The numerical approach has been adopted in this exploratory paper given its computational ease. It is still important in this approach to pick a compactly supported wavelet family. By doing so, the derivative becomes "a linear combination of only a small number of wavelets $\psi'_{j,k}(\tau)$ thus contributing to the numerical stability of the procedure"[50].

Equipped with an understanding of the partial DWT, padding, thresholding and the vaguelette-wavelet decomposition, we are now able to consider the estimation of continuous-time price and quantity functions from high-frequency scanner data.

[48] Compact support means that the wavelet is only non-zero over a finite interval.
[49] Abramovich and Silverman (1998, 7).
[50] Abramovich and Silverman (1998, 7).

## 6. WAVLET-BASED ESTIMATION OF CONTINUOUS-TIME PRICE AND QUANTITY FUNCTIONS

By applying the wavelet methods discussed in section 5, we are now able to estimate continuous-time price and quantity functions using high-frequency scanner data. The steps for this estimation are outlined in the following procedure. The procedure also incorporates the calculation of derivative as it is a similar process.

**Procedure for estimation of continuous-time price and quantity functions and derivatives:**

1. **Choose a suitable mother wavelet.**
2. **Threshold weekly observations to remove any noise/outliers.**
3. **(derivative only) Differentiate the thresholded data series as per the VWD[51].**
4. **Interpolate the discrete observations.**

This section discusses and illustrates each of the steps using the retail scanner data outlined in section 4 of the paper. It is assumed that the weekly average price and quantity observations from the high-frequency scanner data are equally spaced samples of the underlying continuous-time price and quantity functions. As the same procedure applies to price and quantity functions, the steps are illustrated with respect to price functions only.

Recall that the high-frequency scanner data consisted of 418 weekly observations. This is not a dyadic power nor a multiple of a dyadic power. So as previously discussed, it is necessary to pad the data series. The nearest multiple of a dyadic term is 432 hence an extra 14 observations are added. Also, instead of applying the wavelet reconstruction (4), the partial DWT wavelet reconstruction (5) is used with $J_p = 4$.

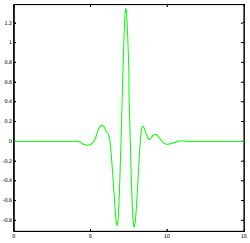
***Step 1 – Choice of Mother Wavelet***

The choice of a suitable mother wavelet is a process of weighing the properties required by the analysis goal of function estimation against the core features of the data. The advantage of choosing a mother wavelet similar in character to the data series is that only a few of dilations and translations

[51] Vaguelette–wavelet decomposition.

of mother wavelet are needed to represent the data. For our analysis, the mother wavelet chosen is the least asymmetric wavelet with length[52] eight (LA(8)) developed by Ingrid Daubechies[53]. It is defined and illustrated below,

$$\psi(t) = e^{-i2\pi(L-1)+i\pi} 2\cos^L(\pi(\frac{1}{2}-t)) \sum_{l=0}^{\frac{L}{2}} \begin{pmatrix} \frac{L}{2}-1+l \\ l \end{pmatrix} \sin^{2l}(\pi t)$$



The LA(8) was chosen because it reasonably captures the characteristics of the data; is continuous; is orthonormal; compactly supported; and its wavelet coefficients are alignable in time[54]. It is important that the same mother wavelet is consistently applied.

The comparison of a sample price series against its wavelet approximations[55] for different mother wavelets, as illustrated Figure 6.1, provides some insight into the selection process. The Haar mother wavelet provides a good fit for the price; it remains constant over weeks with the same prices and accounts for most of the price spikes. It is however inconsistent with our analysis goal of continuous-time function as the Haar wavelet has discontinuities. The other mother wavelets compared are the Daubechies mother wavelet (DB(4))[56] and the LA(8). As the length of the wavelet increases, the better and smoother the approximation is. Ideally, the length should be not too short as it introduces effect and not too large as the chance of boundary effects[57] are increased and there is additional computational burden[58]. It is can be seen from Figure 6.1 that the longer filter of LA(8) provides an improved description over the shorter filter of DB(4). Finally, the LA mother wavelet was preferred because the DB family of wavelets are quite asymmetric.

[52] The length of the mother wavelet relates to the smoothness of the wavelet.
[53] Daubechies (1992). This wavelet is also known as a symmlet.
[54] Percival & Walden (2000, 136).
[55] The approximations are multiresolution approximations at level 1. Refer to Appendix A.2.
[56] Daubechies (1992). The Haar wavelet is actually DB(2). The LA and DB wavelets are different factorisation of the same transfer function, refer to Daubechies (1992) for more information.
[57] As the data is being collected over a finite time interval and not over the entire real line, there needs to be careful treatment of the boundaries. This is beyond the scope of the paper; it will not however discount any conclusions reached in this paper.
[58] Percival & Walden (2000, 135).

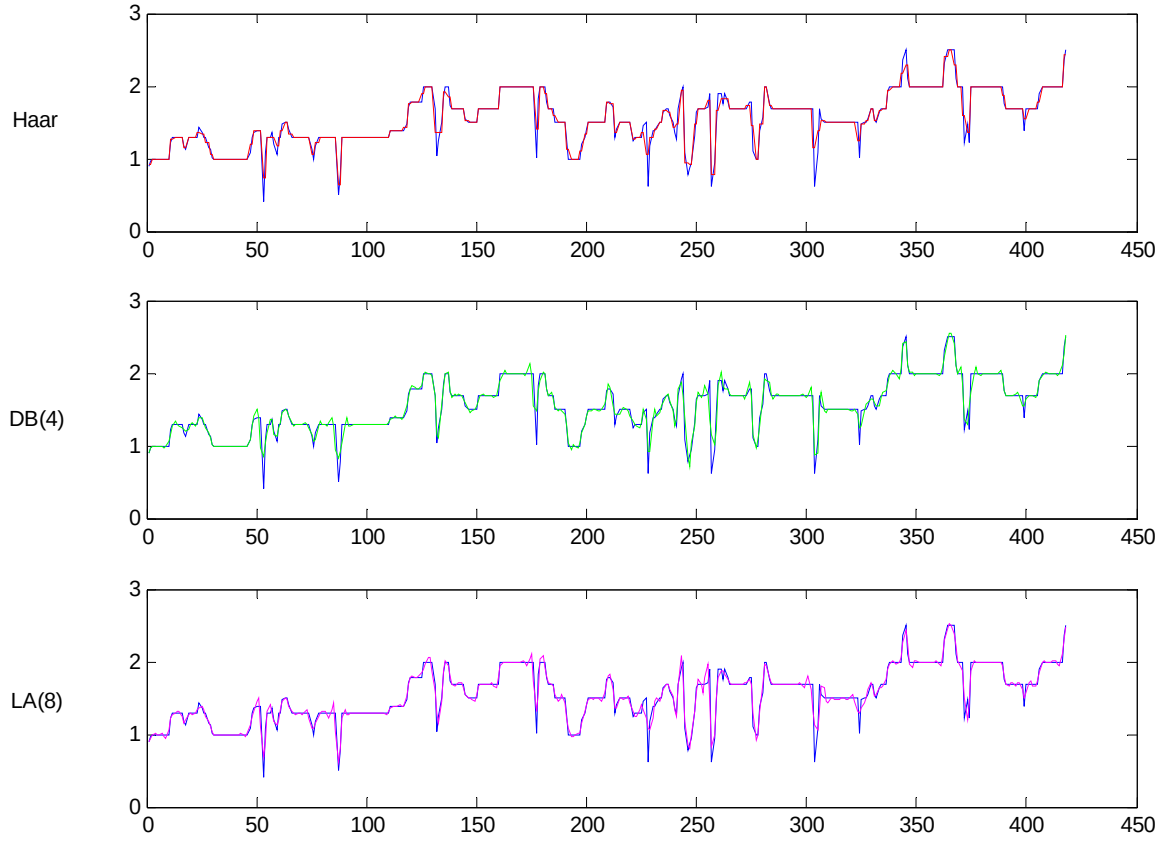



Figure 6.1: Comparison of sample price series against wavelet approximations for different mother wavelets.

***Step 2 – Thresholding***

The process of thresholding involves: getting a wavelet decomposition of the observed data series, applying the soft thresholding scheme at a chosen threshold level λ and then reconstructing the resulting series. This process can be simply expressed for the price function as,

$$p_n^{\tau} = \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \delta_{\lambda}^{S}(d_{j,k}^{n})\psi_{j,k}(\tau) = \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \delta_{\lambda}^{S}(d_{j,k}^{n})2^{j/2}\psi(2^{j}.\tau - k)\,.$$

Using the same sample price series as in step 1, Figure 6.2 shows the effect of soft thresholding at a SURE threshold level.

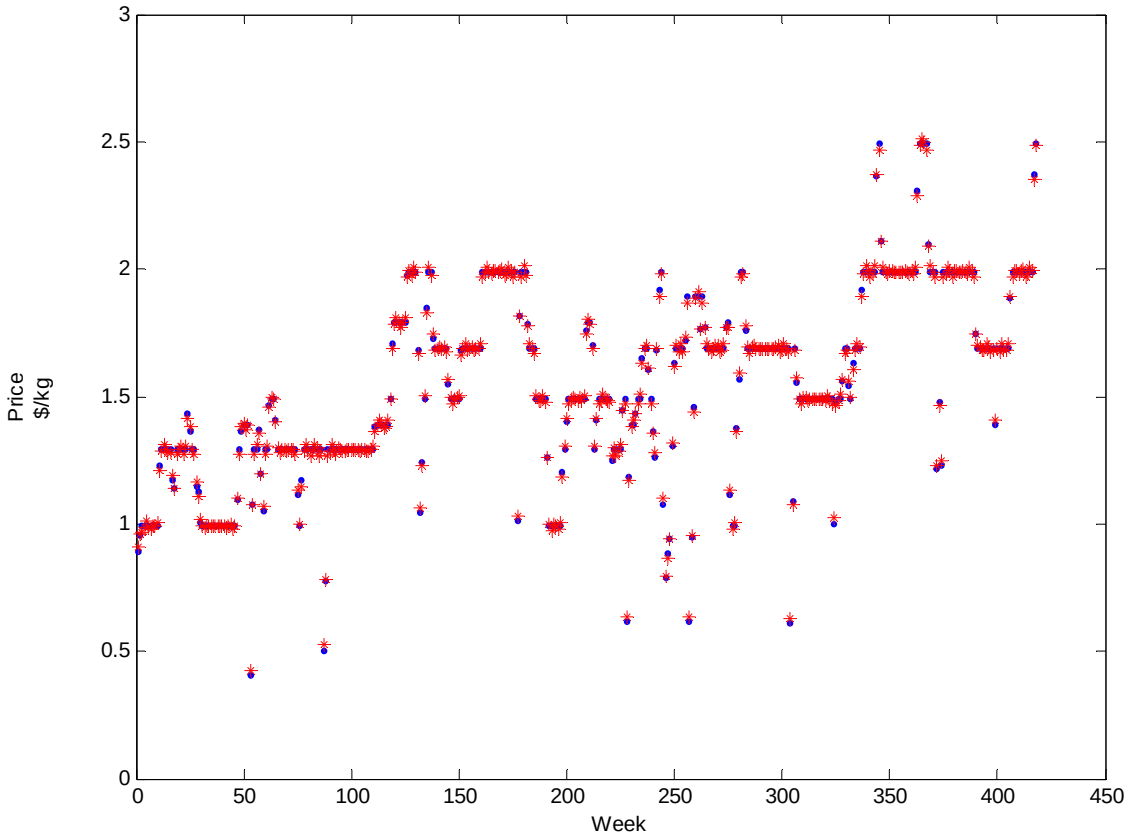



Figure 6.2. Thresholded sample price series using soft thresholding at a SURE threshold level. The blue dots are the original high-frequency price series. The red asterisks are the high-frequency price series after thresholding.

### *Step 3 – (Derivative only) Differentiation*

Using the vaguelette-wavelet decomposition, the derivative of the price function is

$$\frac{d\hat{p}_n(\tau)}{d\tau} = \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \delta_\lambda^S(d_{j,k}^n)\psi'_{j,k}(\tau) = \sum_{j=0}^{J-1} 2^{\frac{3}{2}j} \sum_{k=0}^{2^j-1} \delta_\lambda^S(d_{j,k}^n)\psi'_{j,k}(2^j\tau - k)$$

Figure 6.3 illustrates the derivative of a sample price series in light of the original series using soft thresholding at a SURE threshold.

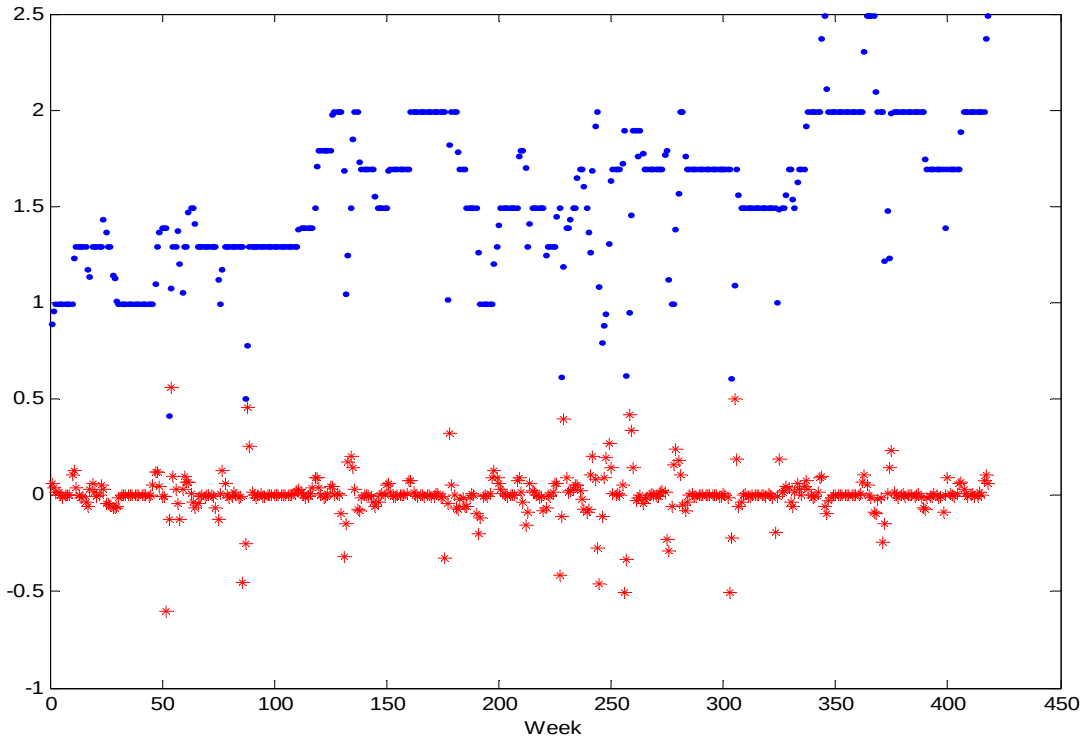



Figure 6.3. Derivative of discrete thresholded function representation of the sample series using soft thresholding at a SURE threshold level. The blue dots are the original high-frequency price series. The red asterisks are the derivative at discrete times.

***Step 4 –Interpolation***

We now have a discrete-time function representation of the weekly observations. In order to obtain a continuous-time representation, we need to interpolate the discrete representation. In doing this, we do not want to impose any excessive structure on the period between observations. Although used for different aspects of wavelet analysis, Sardy et al (1999) used piecewise constant interpolation of observed data to get a continuous-time representation and Hall and Turlach (1997) used local linear interpolation. Given we are using average prices, linear splines are a simple and logical means of interpolation. It may not be realistic to assume that prices change value linearly over a week but it is adequate for the exploratory purposes of this paper. If the frequency of observations increased to daily prices then the effect of this assumption would diminish considerably.

Linear interpolation works by effectively drawing a straight line between neighbouring discrete observations and returning the appropriate point along that line. The *continuous-time price function* is therefore,

$$p_n^{\varepsilon} = \sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi_{j,k}(\eta) + (\tau-\eta)[\sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi_{j,k}(\eta+1) - \sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi_{j,k}(\eta)]$$

such that $0 \le \tau - \eta < 1$ and $\eta = 1,\text{K},418$.

Similarly, the continuous-time price derivative is therefore,

$$\frac{dp_n^{\varepsilon}}{d\tau} = \sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi'_{j,k}(\eta) + (\tau-\eta)[\sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi'_{j,k}(\eta+1) - \sum_{j=0}^{J_p-1}\sum_{k=0}^{2^j-1}\delta_\lambda(d_{j,k}^n)\psi'_{j,k}(\eta)]$$

such that $0 \le \tau - \eta < 1$ and $\eta = 1,\text{K},418$.

Figure 6.4 demonstrates the continuous-time function estimate of the sample prices series using soft thresholding at a SURE threshold level.

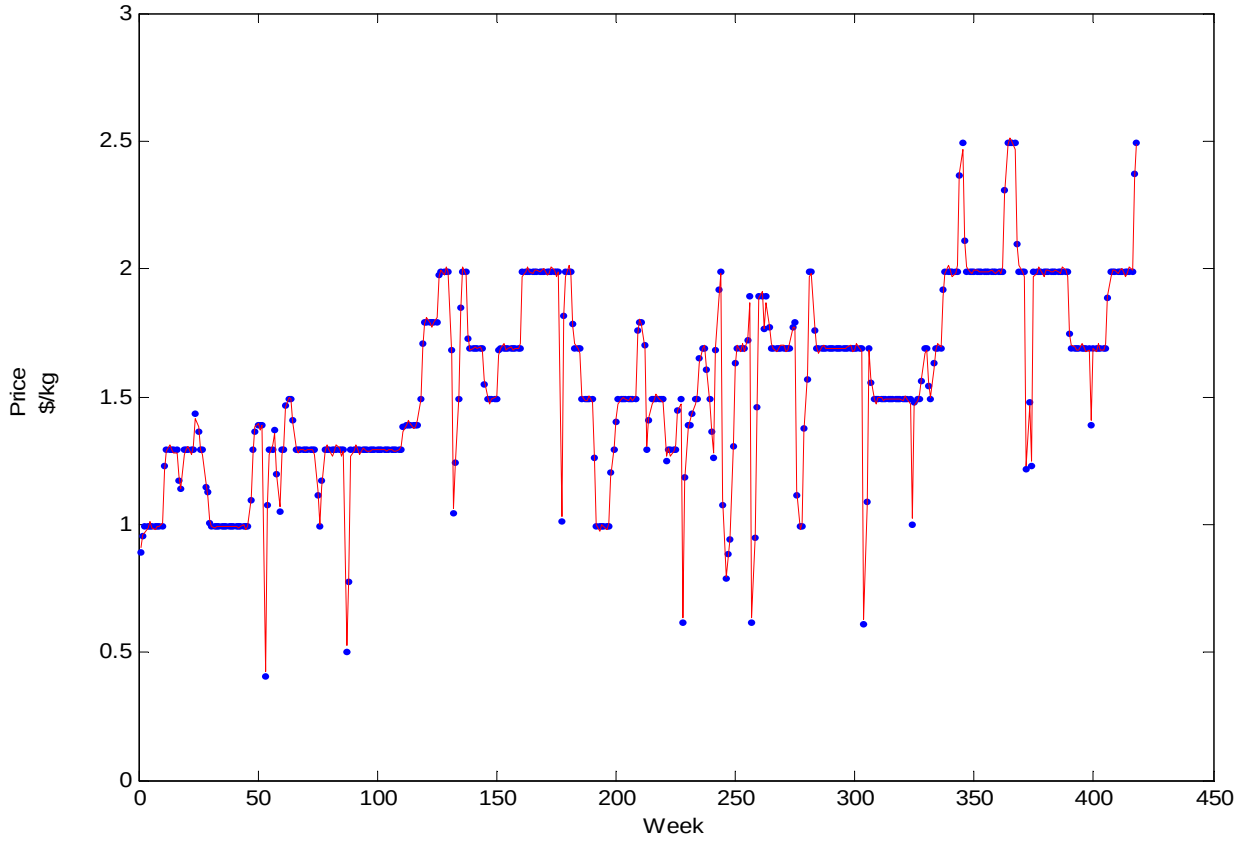



Figure 6.4. Continuous-time function representation of the sample price series using soft thresholding at a SURE threshold level. The blue dots are the original high-frequency price series. The red line is the continuous-time price function.

# 7. EMPIRCAL DIVISIA PRICE AND QUANTITY INDEXES AND DISCRETE-TIME COMPARISONS

## 7.1. Empirical Divisia price and quantity indexes

Given the new ability to construct continuous-time price and quantity functions from high-frequency scanner data, we are now able to compute Divisia price and quantity indexes. Recall that the Divisia price index and the Divisia quantity index are defined as $\ln P_{Divisia}(t',t) \equiv \int_{t'}^{t} \sum_{n=1}^{N} s_n{}^{\tau} \frac{d \ln p_n{}^{\tau}}{d\tau} d\tau$ and $\ln Q_{Divisia}(t',t) \equiv \int_{t'}^{t} \sum_{n=1}^{N} s_n{}^{\tau} \frac{d \ln q_n{}^{\tau}}{d\tau} d\tau$ respectively. The estimation of the Divisia indexes, therefore, involves two key computations – the estimation of the value shares and the estimation of the continuous-time price or quantity logarithmic derivative. These two elements are now discussed in detail.

***Estimation of the value shares*** $s_n{}^{\tau}$

Rather than use observed value shares and create a continuous-time function in a similar manner to that described in section 6, we propose the calculation of value shares from the estimated continuous-time price and quantity functions. This is done using the value share formula $s_n{}^{\tau} \equiv \frac{v_n{}^{\tau}}{v^{\tau}} \equiv \frac{p_n{}^{\tau} q_n{}^{\tau}}{p^{\tau} q^{\tau}}$ where the continuous-time price and quantity functions are as defined in section 6, namely, the continues-time price function is

$$p_n{}^{\varepsilon} = \sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta) + (\tau - \eta)[\sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta+1) - \sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta)]$$

such that $0 \le \tau - \eta < 1$ and $\eta = 1,\text{K},418$ and the continuous time quantity function is

$$q_n{}^{\varepsilon} = \sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta) + (\tau - \eta)[\sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta+1) - \sum_{j=0}^{J_p-1} \sum_{k=0}^{2^j-1} \delta_{\lambda}(d_{j,k}^{n}) \psi_{j,k}(\eta)]$$

such that $0 \le \tau - \eta < 1$ and $\eta = 1,\text{K},418$.

Figure 7.1 illustrates the difference between the observed value shares and the calculated value shares for outlet A. The minimal difference shown in this figure is indicative of our general finding that the calculated value shares are reasonable estimates of the value shares.

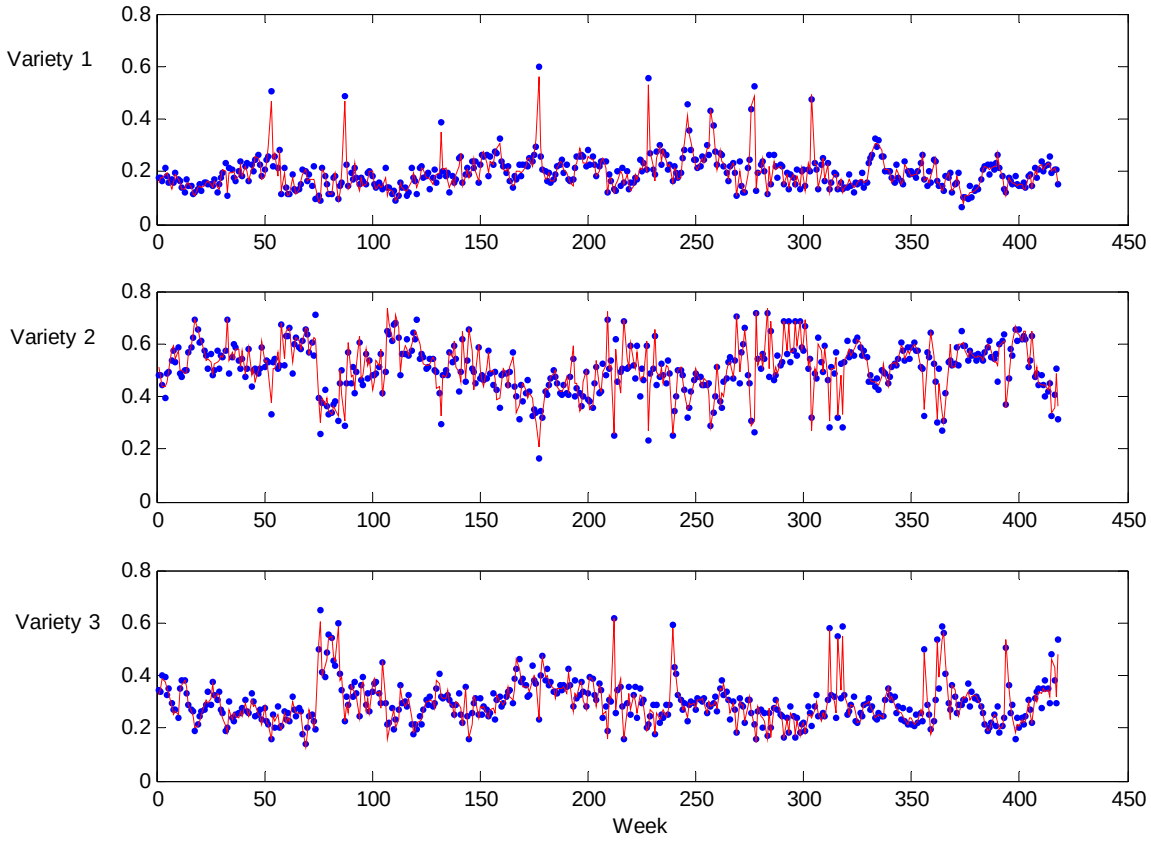



Figure 7.1. The observed value shares (dots) vs. the calculated vale shares at outlet A for the three potato varieties using SURE level soft thresholding (lines).

***Estimation of the logarithmic derivative*** $\frac{d \ln p_n{}^{\tau}}{d\tau}$ ***or*** $\frac{d \ln q_n{}^{\tau}}{d\tau}$

Two options are available for the calculation of the logarithmic derivative: 1) calculate the relevant continuous-time function, differentiate it and then divide the derivative by the original function; or 2) take the natural logarithms of the observed data from which a continuous-time derivative function can be constructed using the method described in section 6. The second option was preferred in this paper as the first option was found on application to introduce a significant element of bias. The bias was further emphasized by the small range of values in the price data.

With the two elements of the Divisia indexes now computed, the calculation of the Divisia indexes is now simply a matter of summing for all commodities the product of the value share and logarithmic derivative and then integrating it over the period of interest. The numerical technique known as the Trapezoidal Rule was used to estimate the integrals. This method approximates an integral with a collection of line segments and integrates under each of the line segments. This approach is consistent with the linear interpolation approach in section 6. Tables 7.1 and 7.2 report the calculated Divisia price and quantity indexes respectively for outlet A across the three potato varieties.

**Table 7.1. DIVISIA PRICE INDEX NUMBERS**
**Outlet A -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^S_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^S_{\lambda_U}$ *(σ =MAD)* | *Divisia* $\delta^S_{\lambda_{MIN}}$ | *Divisia* $\delta^S_{\lambda_{SURE}}$ |
|---|---|---|---|---|
| 1 | 1.3731 | 1.4471 | 1.5035 | 1.5880 |
| 2 | 0.6146 | 0.8476 | 0.8856 | 0.9764 |
| 3 | 1.7645 | 2.4069 | 2.3099 | 2.3574 |
| 4 | 1.1072 | 1.3803 | 1.2999 | 1.2973 |
| 5 | 1.3466 | 1.8569 | 1.7276 | 1.7194 |
| 6 | 0.7085 | 0.8716 | 0.7934 | 0.7737 |
| 7 | 1.0245 | 1.2456 | 1.1530 | 1.1675 |
| 8 | 0.8634 | 1.0192 | 0.9706 | 1.0151 |

**Table 7.2. DIVISIA QUANTITY INDEX NUMBERS**
**Outlet A -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^S_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^S_{\lambda_U}$ *(σ =MAD)* | *Divisia* $\delta^S_{\lambda_{MIN}}$ | *Divisia* $\delta^S_{\lambda_{SURE}}$ |
|---|---|---|---|---|
| 1 | 0.9698 | 1.0630 | 1.1518 | 1.1460 |
| 2 | 1.2570 | 1.0745 | 1.1136 | 1.0085 |
| 3 | 0.7571 | 0.6357 | 0.7043 | 0.6390 |
| 4 | 1.4049 | 1.4019 | 1.6316 | 1.4787 |
| 5 | 1.5336 | 1.3669 | 1.6196 | 1.4539 |
| 6 | 1.6515 | 1.7032 | 2.2242 | 2.0957 |
| 7 | 1.6457 | 1.8092 | 2.3303 | 2.1012 |
| 8 | 1.3250 | 1.4549 | 1.7531 | 1.5796 |

As expected, the Divisia indexes calculated using the minimax and SURE thresholds are relatively similar with the universal threshold Divisia indexes generally lower.

It is also interesting to compare the observed value ratio against the product of the Divisia price and quantity indexes. Figure 7.2 illustrates this comparison at outlet A. The figure indicates that the Divisia indexes provide a reasonable fit though a slight degree of upward bias is evident.

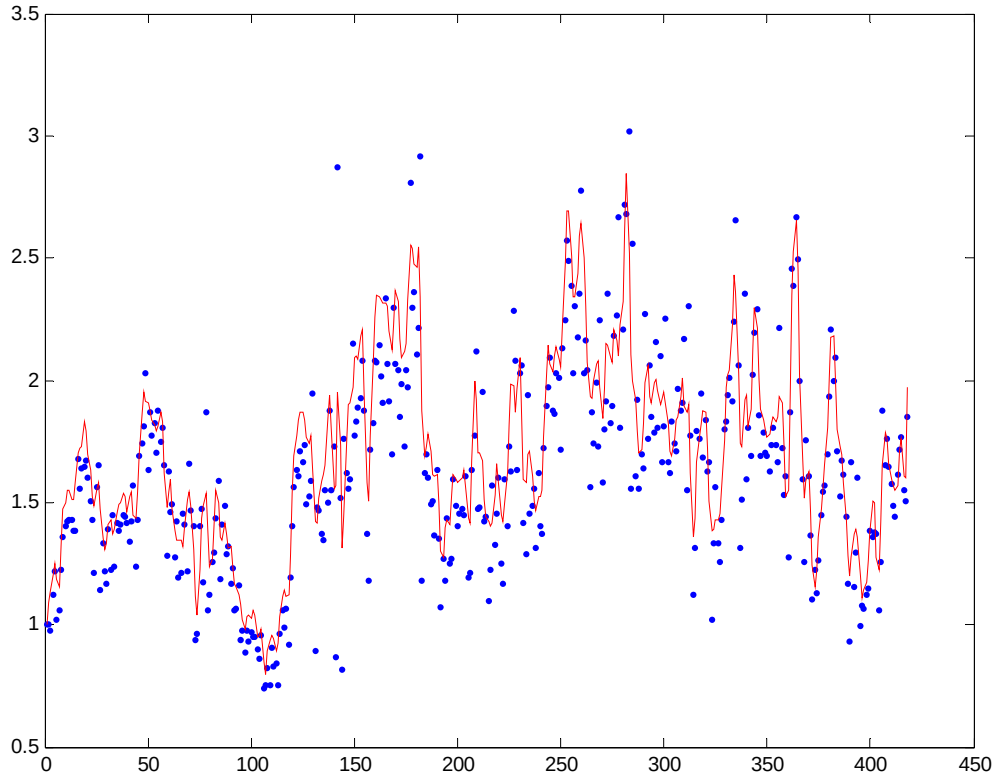



Figure 7.2. The observed value ratio (in blue dots) vs. the product of the Divisia price index and Divisia quantity index at outlet A (across the three potato varieties) using SURE thresholding (red line).

## 7.2. Comparison of calculated Divisia indexes to common discrete-time approximations

Two types of indexes are used to compare the calculated Divisia indexes to common discrete-time Divisia index approximations– fixed base indexes and chained indexes. The base period price is the price at period 0 and the base period quantity was the quantity sold at period 0.

Tables 7.3 and 7.4 report the results of a fixed base price and quantity index comparison for outlet A respectively. These indexes are computed using the formulae listed in Table 3.1. The tables are also illustrated graphically.

**Table 7.3. COMPARISON OF FIXED BASE PRICE INDEX NUMBERS**
**Outlet A**
(Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^{S}_{\lambda_U}$ *(σ=MAD)* | *Divisia* $\delta^{S}_{\lambda_{MIN}}$ | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Törnqvist-Theil* | *Sato-Vartia* | *Fixed-base Laspeyres* | *Fixed-base Paasche* |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.3731 | 1.4471 | 1.5035 | 1.5880 | 1.3813 | 1.4200 | 2.1639 | 0.7633 |
| 2 | 0.6146 | 0.8476 | 0.8856 | 0.9764 | 1.1139 | 1.1459 | 1.3277 | 0.9657 |
| 3 | 1.7645 | 2.4069 | 2.3099 | 2.3574 | 2.4383 | 2.5839 | 3.0332 | 2.2127 |
| 4 | 1.1072 | 1.3803 | 1.2999 | 1.2973 | 1.1874 | 1.1865 | 1.2777 | 1.1149 |
| 5 | 1.3466 | 1.8569 | 1.7276 | 1.7194 | 2.5094 | 2.5046 | 2.6297 | 2.3924 |
| 6 | 0.7085 | 0.8716 | 0.7934 | 0.7737 | 1.3439 | 1.3210 | 1.3920 | 1.2427 |
| 7 | 1.0245 | 1.2456 | 1.1530 | 1.1675 | 2.3101 | 2.1785 | 3.1485 | 1.4847 |
| 8 | 0.8634 | 1.0192 | 0.9706 | 1.0151 | 2.8918 | 2.8336 | 2.9239 | 2.7446 |

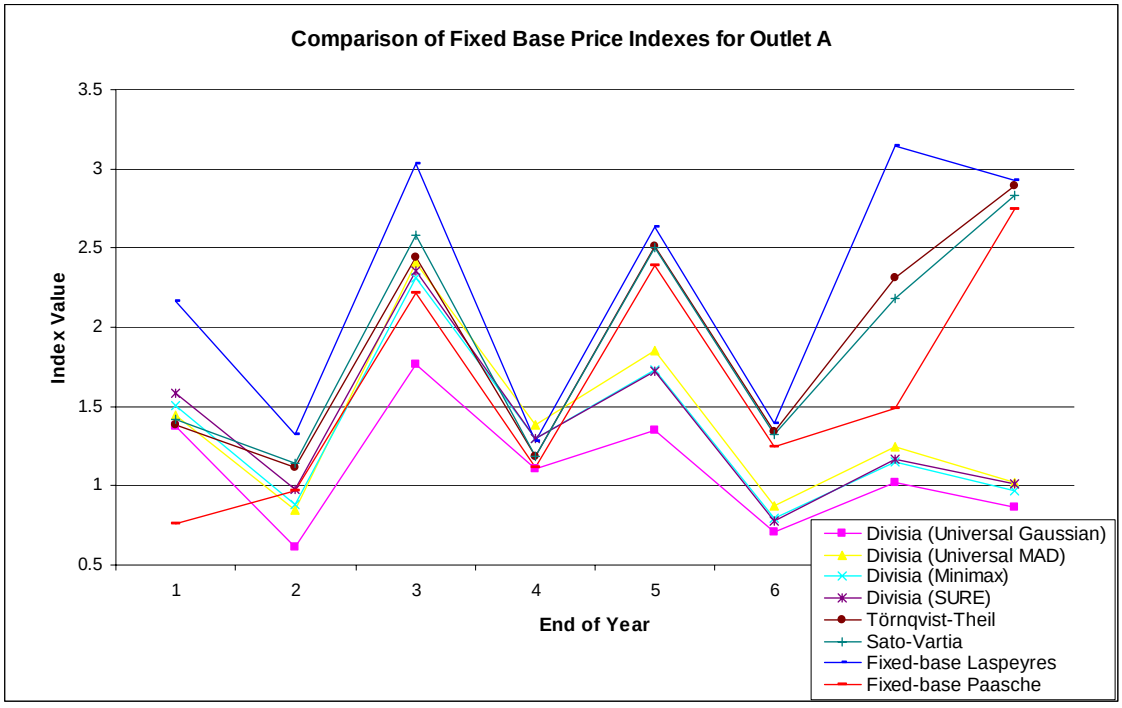
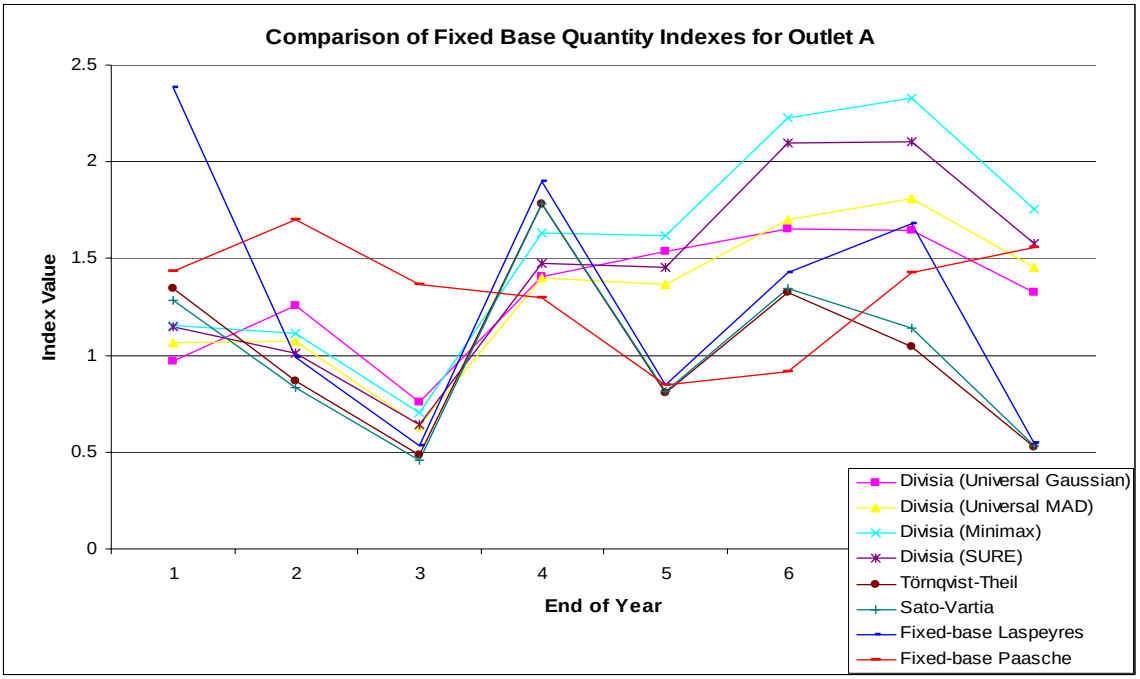


**Table 7.4. COMPARISON OF FIXED BASE QUANTITY INDEX NUMBERS**
**Outlet A**
(Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^S_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^S_{\lambda_U}$ *(σ =MAD)* | *Divisia* $\delta^S_{\lambda_{MIN}}$ | *Divisia* $\delta^S_{\lambda_{SURE}}$ | *Törnqvist-Theil* | *Sato-Vartia* | *Fixed-base Laspeyres* | *Fixed-base Paasche* |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.9698 | 1.0630 | 1.1518 | 1.1460 | 1.3462 | 1.2832 | 2.3872 | 1.4347 |
| 2 | 1.2570 | 1.0745 | 1.1136 | 1.0085 | 0.8676 | 0.8321 | 0.9873 | 1.6999 |
| 3 | 0.7571 | 0.6357 | 0.7043 | 0.6390 | 0.4817 | 0.4557 | 0.5321 | 1.3637 |
| 4 | 1.4049 | 1.4019 | 1.6316 | 1.4787 | 1.7849 | 1.7827 | 1.8972 | 1.2994 |
| 5 | 1.5336 | 1.3669 | 1.6196 | 1.4539 | 0.8054 | 0.8105 | 0.8485 | 0.8452 |
| 6 | 1.6515 | 1.7032 | 2.2242 | 2.0957 | 1.3249 | 1.3431 | 1.4278 | 0.9170 |
| 7 | 1.6457 | 1.8092 | 2.3303 | 2.1012 | 1.0445 | 1.1439 | 1.6785 | 1.4270 |
| 8 | 1.3250 | 1.4549 | 1.7531 | 1.5796 | 0.5238 | 0.5303 | 0.5475 | 1.5581 |




The Divisia indexes differ quite substantially from the more common discrete-time fixed-base index numbers. This is somewhat expected as the Divisia indexes incorporate all price and quantity information whereas the discrete-time index numbers are only looking at the end-of-year points. This explains why the Divisia price indexes in the last year move downward while at the same time the Törnqvist-Theil, Paasche and Sato-Vartia indexes rise.

Tables 7.5 and 7.6 report the results of the chained price index comparison and quantity index comparison for outlet A respectively. The chained discrete-time approximation indexes are computed using the formula listed in the headings. The tables are also illustrated graphically.

**Table 7.5. COMPARISON OF CHAINED PRICE INDEX NUMBERS**
**Outlet A -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^S_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^S_{\lambda_U}$ *(σ =MAD)* | *Divisia* $\delta^S_{\lambda_{MIN}}$ | *Divisia* $\delta^S_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} P_{TT}(t,t-1)$ where $P_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} P_F(t,t-1)$ where $P_F(1,0)=1$ |
|---|---|---|---|---|---|---|
| 1 | 1.3731 | 1.4471 | 1.5035 | 1.5880 | 1.2573 | 1.2353 |
| 2 | 0.6146 | 0.8476 | 0.8856 | 0.9764 | 0.9521 | 0.9255 |
| 3 | 1.7645 | 2.4069 | 2.3099 | 2.3574 | 2.1074 | 1.9687 |
| 4 | 1.1072 | 1.3803 | 1.2999 | 1.2973 | 0.8293 | 0.7990 |
| 5 | 1.3466 | 1.8569 | 1.7276 | 1.7194 | 1.5472 | 1.4503 |
| 6 | 0.7085 | 0.8716 | 0.7934 | 0.7737 | 0.5929 | 0.5485 |
| 7 | 1.0245 | 1.2456 | 1.1530 | 1.1675 | 1.0785 | 0.9868 |
| 8 | 0.8634 | 1.0192 | 0.9706 | 1.0151 | 1.0904 | 0.9962 |

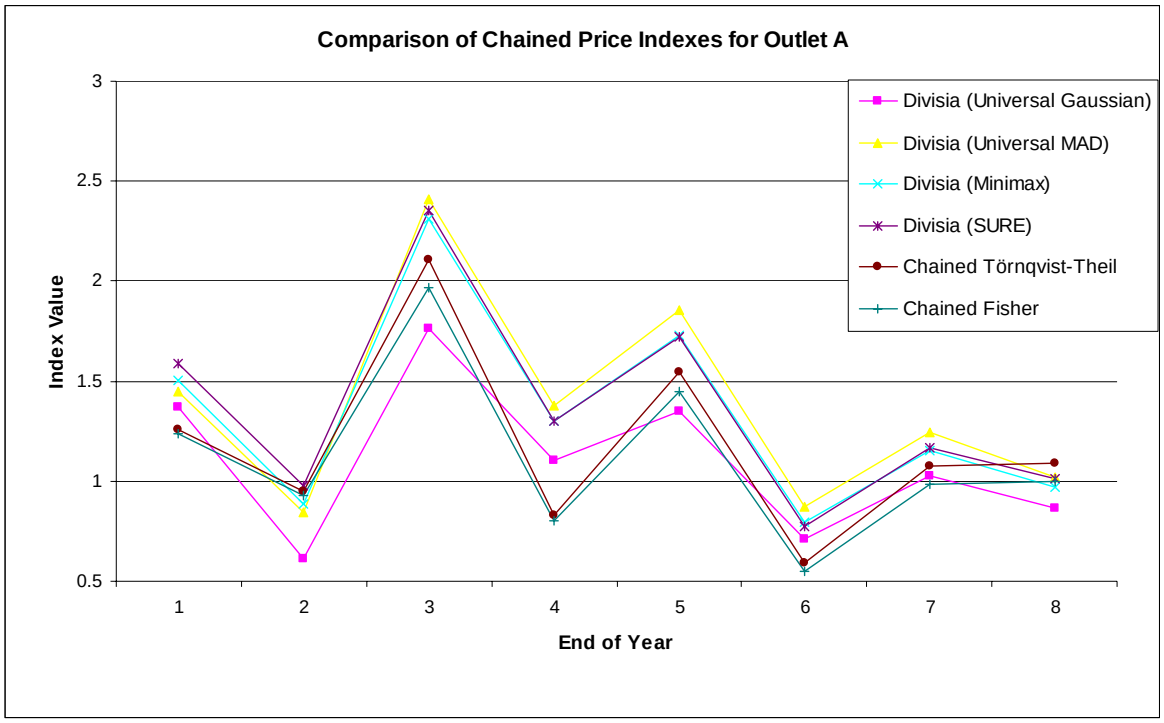
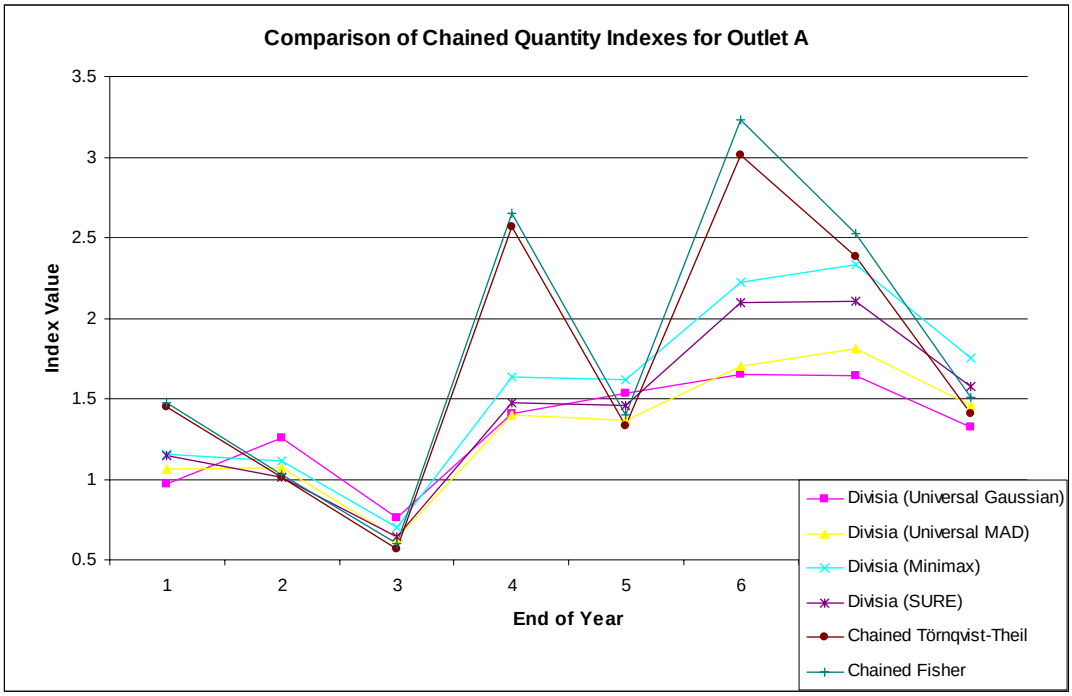


**Table 7.6. COMPARISON OF CHAINED QUANTITY INDEX NUMBERS**
**Outlet A -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_U}$ *(σ =1)* | *Divisia* $\delta^{S}_{\lambda_U}$ *(σ =MAD)* | *Divisia* $\delta^{S}_{\lambda_{MIN}}$ | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} Q_{TT}(t,t-1)$ where $Q_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} Q_{F}(t,t-1)$ where $Q_{F}(1,0)=1$ |
|---|---|---|---|---|---|---|
| 1 | 0.9698 | 1.0630 | 1.1518 | 1.1460 | 1.4531 | 1.4751 |
| 2 | 1.2570 | 1.0745 | 1.1136 | 1.0085 | 1.0120 | 1.0302 |
| 3 | 0.7571 | 0.6357 | 0.7043 | 0.6390 | 0.5669 | 0.5981 |
| 4 | 1.4049 | 1.4019 | 1.6316 | 1.4787 | 2.5637 | 2.6473 |
| 5 | 1.5336 | 1.3669 | 1.6196 | 1.4539 | 1.3293 | 1.3996 |
| 6 | 1.6515 | 1.7032 | 2.2242 | 2.0957 | 3.0133 | 3.2348 |
| 7 | 1.6457 | 1.8092 | 2.3303 | 2.1012 | 2.3839 | 2.5253 |
| 8 | 1.3250 | 1.4549 | 1.7531 | 1.5796 | 1.4092 | 1.5085 |




From the above calculations, the Divisia price indexes appear to share significant similarities with the chained discrete-time Törnqvist-Theil and Fisher indexes. While the same degree of similarity is not reflected in the quantity index comparison, this can be explained by a few very large quantity spikes during the periods of divergence. Unlike the approaches of the chained discrete-time Törnqvist-Theil

and Fisher indexes, the quantity spikes are significantly reduced in the wavelet-based Divisia approach.

These comparisons of wavelet-based Divisia indexes and discrete-time approximations for outlet A provide some interesting results especially in case of the chained index comparisons. The similarity of the Divisia indexes and chained indexes offers some support for the use of chained indexes. Comparisons of chained price and quantity indexes for outlet B and C are shown in Appendix B.1. These comparisons also provide some support for the use of chained indexes.

Therefore, the examination on how the calculated Divisia indexes compare to price and quantity indexes computed using more conventional statistical index number methods has established the viability of wavelet-based Divisia indexes for this particular subset of scanner data. The similarity of the Divisia indexes and chained discrete-time indexes is particularly exciting as it provides some support for the use of chained indexes in practice.

## 8. CONCLUSIONS AND FUTURE RESEARCH

The main purpose of this paper was to explore the potential of wavelet methods to estimate continuous-time functions from high-frequency scanner data on prices and quantities. The approach here deviates from current approaches to the construction of price indexes and attempts to exploit the almost continuous nature of observed price and quantity data. The paper has shown successfully how wavelet representations can be used to model price and quantity high-frequency data as a function of time and thereby established the necessary apparatus to calculate Divisia price and quantity indexes. The empirical example based on price data on potatoes shows the feasibility of this approach and provide some support for the use of chained indexes.

There are still a number of outstanding issues that need further consideration. One such issue is the detection and estimation of a seasonal component in the high-frequency scanner data. Wavelet methods present a new means of identifying the presence of seasonal patterns in the price and quantity data – this could be of significance in the context of fruits and vegetables. Another issue is the consideration of Divisia indexes in light of different base periods like, for instance, the use of the modal price as the base period price as in Feenstra and Shapiro (2003). Future work should also further consider the construction of Divisia price and quantity indexes across outlets, specifically the averaging methods across outlets. Another area for future work is the use of wavelet methods for the prediction of missing data. A final issue relates to technical wavelet analysis aspects. In particular, further work is necessary in establishing procedures for selecting appropriate wavelets, thresholds and interpolation techniques necessary for successful application of wavelet methods in the construction of price and quantity indexes using scanner data.

## APPENDIX A

### A.1 Value of wavelets over other methods particularly Fourier analysis

Wavelets are valuable because of their good time-frequency localisation, ability to handle non-stationary data, ease of reproducibility and simplicity of form. Compared to Fourier analysis, wavelet analysis can capture features that are local in both time and frequency, and can automatically adapt to the amount of time and frequency localisation in the time series. Consider a pianist playing a series of chords[59], wavelet analysis can recapture the timing of the chords and the chords themselves. Fourier analysis, in comparison, can at best recapture the chords but not their order of play. In its classical form, Fourier analysis cannot even determine which notes belong to which chords. Wavelet analysis also has a real advantage over traditional time series methods because it can handle non-stationary or time-varying phenomena particularly volatile data as well as stationary processes. Furthermore, the simplicity of the wavelet form and the need for only a few assumptions means wavelet methods are computationally faster than standard statistical methods. Finally, the ease with which results can be reproduced using wavelet methods is an important practical feature. Consequently, wavelet methods are a very appealing means of representing time series.

### A.2 Multiresolution analysis

The fundamental idea of multiresolution analysis (MRA) is to represent a time series as a limit of successive approximations, each of which is a smoother version of the time series. These successive approximations correspond to different levels of resolution hence the name MRA.

The notion of the MRA is best demonstrated using the DWT. Consider the DWT in terms of matrices. The $N$ dimensional vector of DWT coefficients $\mathbf{W}$ is obtained via $\mathbf{W} = W\mathbf{X}$ where $\mathbf{X}$ is the $N$ dimensional vector of the observations and $W$ is a real-valued $N \times N$ matrix composed of wavelet coefficient vectors $\mathbf{W}_1, \ldots, \mathbf{W}_J$ and a vector of scaling coefficients $\mathbf{V}_J$[60]. The DWT coefficients $\mathbf{W}$ and the matrix $W$ can be partitioned so $\mathbf{W}_J = W_J\mathbf{X}$ and $\mathbf{V}_J = V_J\mathbf{X}$. Therefore, the expression $\mathbf{W} = W\mathbf{X}$ is equivalently,

[59] Musical analogy based on Ogden (1997,1)
[60] The DWT is usually calculated in practice without actually exhibiting the matrix $W$. This is accomplished using the fast filtering algorithms described in Appendix A.3.

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathrm{M} \\ \mathbf{W}_J \\ \mathbf{V}_J \end{bmatrix} = \begin{bmatrix} W_1\mathbf{X} \\ W_2\mathbf{X} \\ \mathrm{M} \\ W_J\mathbf{X} \\ V_J\mathbf{X} \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \mathrm{M} \\ W_J \\ V_J \end{bmatrix} \mathbf{X} = W\mathbf{X}$$

The matrix $W$ is orthonormal, that is, $W^TW = I_N$[61] and hence, the DWT is an orthonormal transform. Consequently, the 'best' approximation in a least squares sense for the time series is formed by $\mathbf{X} = W^T\mathbf{W}$. This is the MRA – the additive decomposition of a data series in terms of $N$ dimensional vectors, each of which can be associated with a particular resolution/scale. $\mathbf{X} = W^T\mathbf{W}$ can be equivalently expressed as

$$\mathbf{X} = \sum_{j=1}^{J} W_j^T\mathbf{W_J} + V_J^T\mathbf{V_J} \equiv \sum_{j=1}^{J} D_j + A_J$$

$D_j \equiv W_j^T\mathbf{W}_j$ is the $j$-th wavelet detail ($j$=1,…,$J$) which refers to the variations in the time series at scale $\tau_j$=$2^{j-1}$ and $A_J \equiv V_J^T\mathbf{V}_J$ equals the sample mean of the observations (the global view)[62].

## A.3 Practical Wavelet Decomposition and Reconstruction

While conceptually wavelet transforms use dilations and translations of the mother wavelet, fast computational algorithms in practice use a combination of mother wavelets $\psi(t)$ and scaling functions $\varphi(t)$. The scaling function is used to describe the averages of the original data across various scales[63]. The mother wavelet is used to describe the differences between the spaces spanned by the various scales of the scaling function.

---

[61] Recall from Note 1 that, discrete wavelet functions must be orthogonal to its even shifts. This means that the wavelet coefficients cannot interact with one another.

[62] Another graphical feature sometimes used is the wavelet smooth $S_j = \sum_{k=j+1}^{J+1} D_j$ which shows the cumulative sum of the variations in the time series at different scales.

[63] A scaling function is also known as a father wavelet. Most but not all wavelet functions have an associated scaling function.

The computational technique derives wavelet and scaling coefficients[64] from wavelet and scaling functions. The derivation of the scaling coefficients $g_l$ and wavelet coefficients $h_l$ is achieved through the scaling equation, $\varphi(t) = \sum_l g_l \sqrt{2}\varphi(2t - l), l \in \boldsymbol{Z}$ and wavelet equation, $\psi(t) = \sum_l h_l \sqrt{2}\psi(2t - l), l \in \boldsymbol{Z}$ respectively. These equations and consequently the scaling and wavelet coefficients are governed by the quadrature mirror relationship $g_l = (-1)^l h_{L-l}$ for $l = 1,..., L$. This ensures that the scaling coefficients reflect the long-term variations in the data (low pass filters) and wavelet coefficients reflect short-term changes (high pass filters) in the data.

[64] Also known as high-pass filter coefficients and low-pass filter coefficients respectively.

## APPENDIX B

### B.1 Comparison of chained index numbers

**Table B.1. COMPARISON OF CHAINED PRICE INDEX NUMBERS**
**Outlet B -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} P_{TT}(t,t-1)$ where $P_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} P_{F}(t,t-1)$ where $P_{F}(1,0)=1$ |
|---|---|---|---|
| 1 | 1.4718 | 1.5952 | 1.6004 |
| 2 | 0.8735 | 1.0237 | 0.9989 |
| 3 | 1.2021 | 1.0655 | 1.0088 |
| 4 | 0.8771 | 1.0069 | 0.8740 |
| 5 | 1.3273 | 1.3449 | 1.1424 |
| 6 | 0.6253 | 0.5559 | 0.4190 |
| 7 | 1.0022 | 1.1099 | 0.7304 |
| 8 | 0.7408 | 0.8892 | 0.5605 |

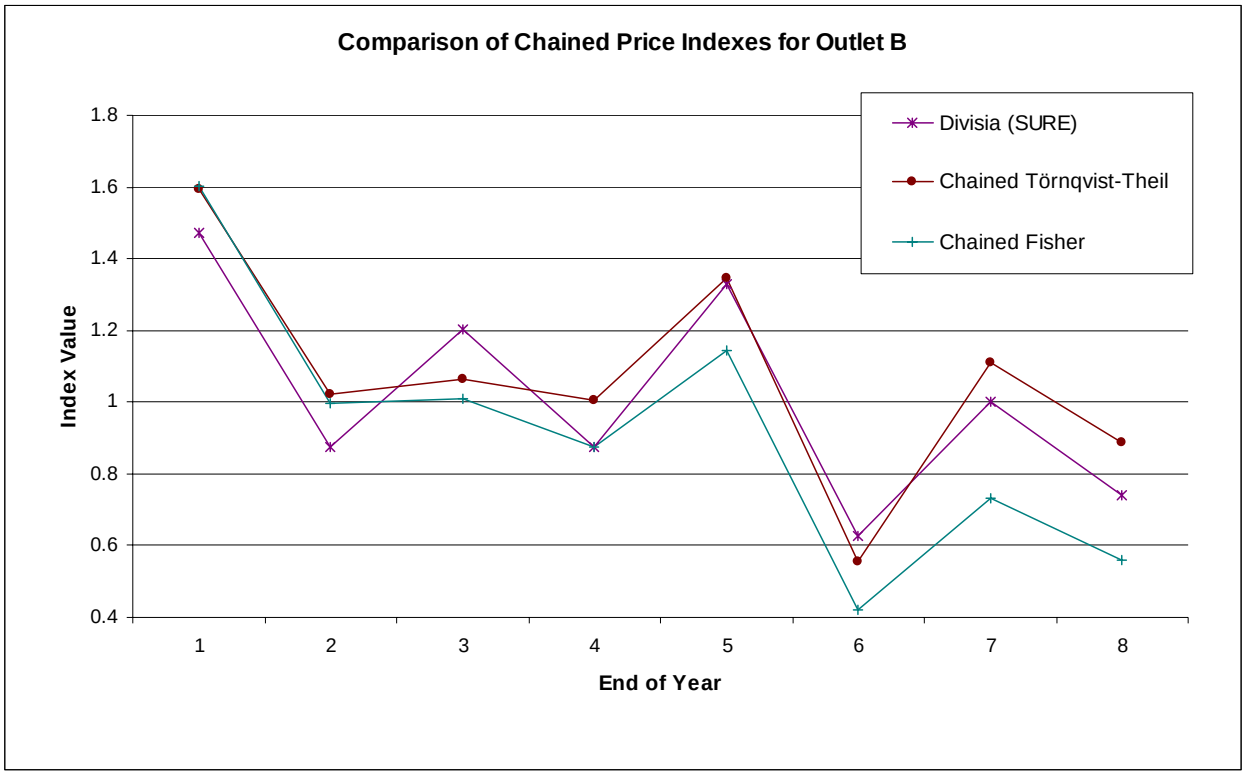


**Table B.2. COMPARISON OF CHAINED QUANTITY INDEX NUMBERS**
**Outlet B -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} Q_{TT}(t,t-1)$ where $Q_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} Q_{F}(t,t-1)$ where $Q_{F}(1,0)=1$ |
|---|---|---|---|
| 1 | 0.8802 | 0.7751 | 0.7725 |
| 2 | 1.0794 | 0.9427 | 0.9756 |
| 3 | 1.1448 | 1.4625 | 1.5630 |
| 4 | 1.3795 | 1.0864 | 1.2760 |
| 5 | 1.1768 | 1.1719 | 1.4044 |
| 6 | 2.6762 | 2.8901 | 3.8336 |
| 7 | 2.3134 | 1.7825 | 2.6587 |
| 8 | 3.3853 | 1.8748 | 2.9497 |

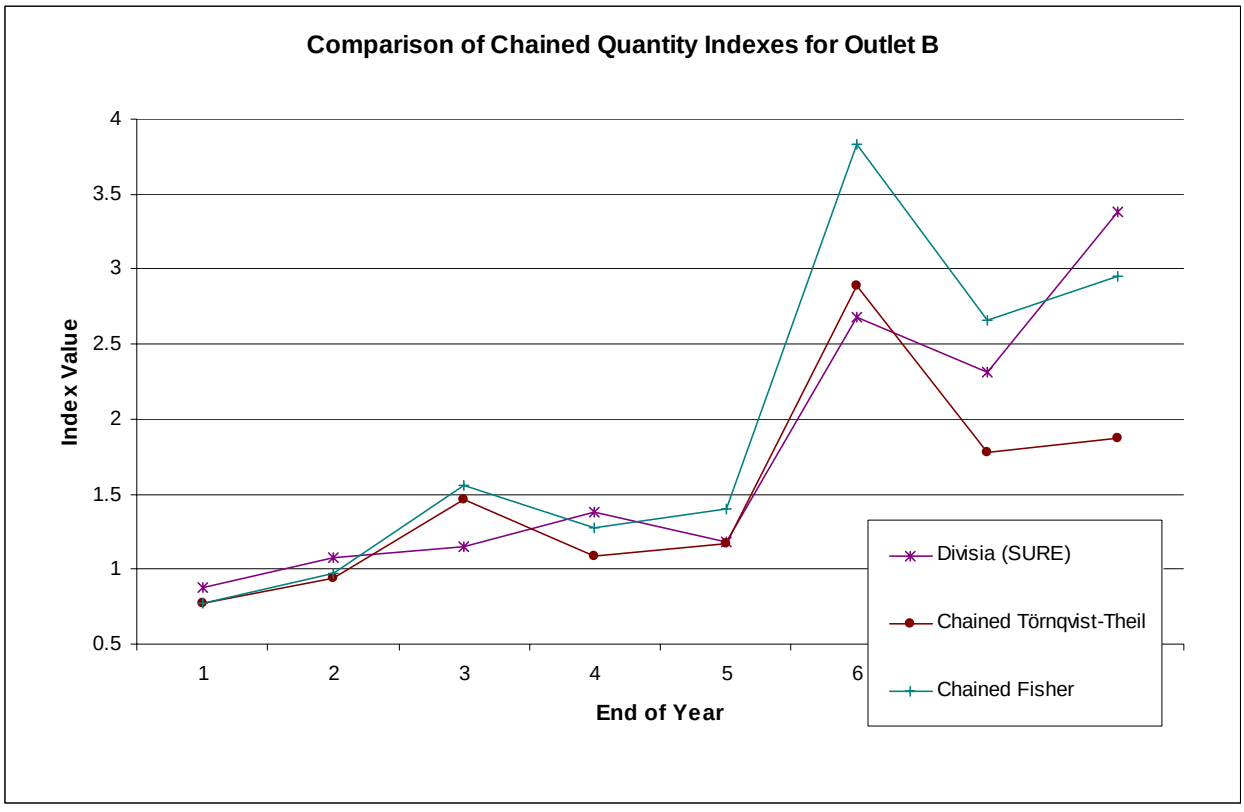


**Table B.3. COMPARISON OF CHAINED PRICE INDEX NUMBERS**
**Outlet C -** (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} P_{TT}(t,t-1)$ where $P_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} P_{F}(t,t-1)$ where $P_{F}(1,0)=1$ |
|---|---|---|---|
| 1 | 1.2682 | 1.1777 | 1.1704 |
| 2 | 0.6464 | 0.7004 | 0.6894 |
| 3 | 0.7828 | 0.9246 | 0.8785 |
| 4 | 0.5388 | 0.6227 | 0.5919 |
| 5 | 0.7459 | 0.9116 | 0.8182 |
| 6 | 0.4951 | 0.5770 | 0.5063 |
| 7 | 0.6907 | 0.8115 | 0.6821 |
| 8 | 0.7076 | 0.8728 | 0.7187 |

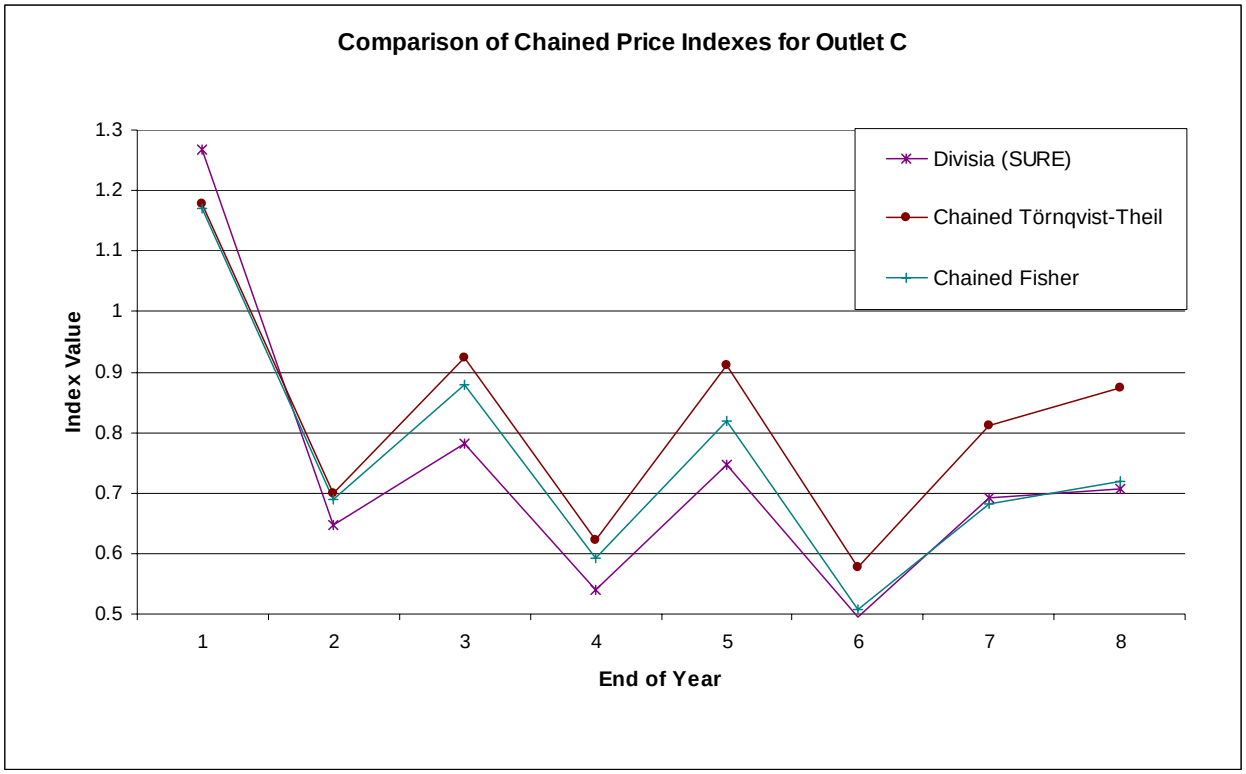
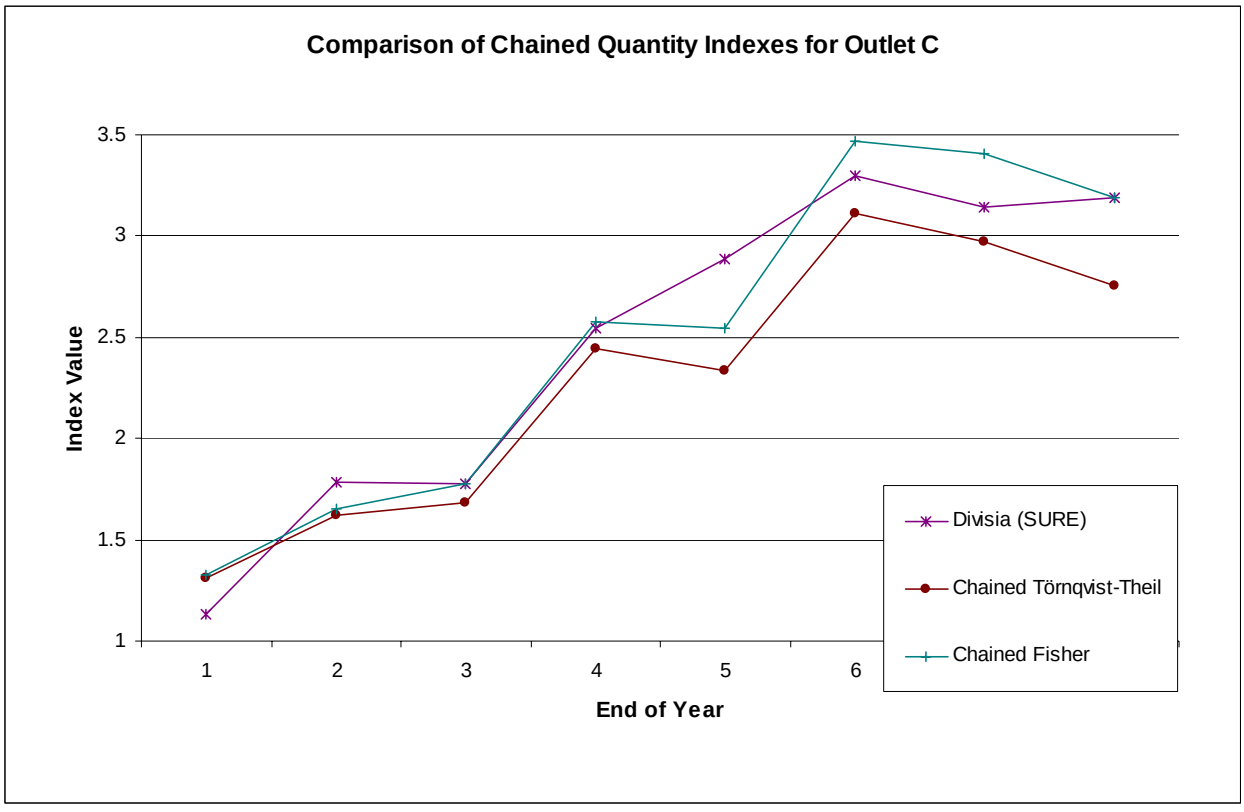


**Table B.4. COMPARISON OF CHAINED QUANTITY INDEX NUMBERS**
**Outlet C** - (Base period: Period 0)

| End of Year (t) | *Divisia* $\delta^{S}_{\lambda_{SURE}}$ | *Chained Törnqvist-Theil* $\prod_{t=1}^{T} Q_{TT}(t,t-1)$ where $Q_{TT}(1,0)=1$ | *Chained Fisher* $\prod_{t=1}^{T} Q_{F}(t,t-1)$ where $Q_{F}(1,0)=1$ |
|---|---|---|---|
| 1 | 1.1314 | 1.3088 | 1.3248 |
| 2 | 1.7859 | 1.6190 | 1.6542 |
| 3 | 1.7794 | 1.6829 | 1.7750 |
| 4 | 2.5421 | 2.4437 | 2.5736 |
| 5 | 2.8840 | 2.3346 | 2.5467 |
| 6 | 3.2973 | 3.1105 | 3.4666 |
| 7 | 3.1438 | 2.9703 | 3.4052 |
| 8 | 3.1896 | 2.7571 | 3.1892 |



## REFERENCES

F. Abramovich and B.W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85: 115-129, 1998.

B.M. Balk. *Divisia price and quantity indexes*: 75 years after. Draft Research Paper, Statistics Netherlands, Voorburg, 2000.

W.A. Barnett and A. Serletis (eds.). *The Theory of Monetary Aggregation*. North Holland, Amsterdam, 2000.

A.R. Bergstrom. Survey of Continuous Time Econometrics. 3-25, 1996. In W.A. Barnett, G. Gandolfo and C. Hillinger (eds). *Dynamic Disequilibrium Modeling: Theory and Applications*. Cambridge University Press, New York, 1996.

I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM, Philadelphia, 1992.

W.E. Diewert. Exact and Superlative Index Numbers. *Journal of Econometrics*, 4: 114-145, 1976.

W.E. Diewert. The Early History of Price Index Research. 33-65, 1993. In W.E. Diewert and A.O. Nakamura (eds). *Essays in Index Number Theory*. North-Holland, Amsterdam, 1993.

W.E. Diewert. *The Consumer Price Index and index number theory: a survey*. Discussion Paper 01-02, University of British Columbia, Department of Economics, February 2001.

F. Divisia. L' Indice Mon_etaire et la Theorie de la Monnaie. *Revue d' Economie Politique,* 39: 842-861; 980-1008 and 1121-1151, 1925.

D.L. Donoho. *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*. Technical Report 403, Stanford University, Department of Statistics, January 1992.

D.L. Donoho and I.M. Johnston. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425-455, 1994.

D.L. Donoho and I.M. Johnston. Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association*, 90: 1200-1224, 1995.

R.C. Feenstra and M.D. Shapiro. High-Frequency Substitution and the Measurement of Price Indexes. 123-146, 2003. In R.C. Feenstra and M.D. Shapiro (eds). *Scanner Data and Price Indexes*. University of Chicago Press, Chicago, 2003.

R. Frisch. Annual Survey of General Economic Theory: The Problem of Index Numbers. *Econometrica* 4: 1-38, 1936.

G. Gandolfo. *Economic Dynamics*. Springer-Verlag, Berlin, 1997.

P. Hall and B.A. Turlack. Interpolation Methods for Nonlinear Wavelet Regression with IRregularl Spaced Design. *Annals of Statistics*, 25(5): 1912-12925, 1997.

C.R. Hulten. Divisia Index Numbers. *Econometrica*, 41: 1017-1025, 1973.

E.D. Kolaczyk. *Wavelet Methods for the Inversion of Certain Homogeneous Linear Operators in the Presence of Noisy Data*. Ph.D. thesis, Stanford University, 1994.

R.T. Ogden. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston, 1997.

D.B. Percival and A.T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, London, 2000.

S. Sardy, D.B. Percival, A.G. Bruce., H. Hao and W.Stuetzle. *Wavelet Shrinkage for Unequally Spaced Data*, 9: 65-75, 1999.

K. Sato. The Ideal Log-Change Index Number. *The Review of Economics and Statistics*, 58: 223-228, 1976.

C.M. Schut. *Using Scanerdata to Calculate Consumer Price Indices*. Statistics Netherlands, Voorburg, 2002

C.M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics*, 9: 1135-1151, 1981.

H. Theil. *Economics and Information Theory*. North-Holland, Amsterdam, 1967.

L. Törnqvist. The Bank of Finland's Consumption Price Index. *Bank of Finland Monthly Bulletin*, 10: 1-8, 1936

P.K. Trivedi. Some Discrete Approximations to Divisia Integral Indices. *International Economic Review*, 22: 71-77, 1981.

Y. O. Vartia. *Relative Changes and Economic Indices*. Licensiate Thesis, University of Helsinki, 1974.

Y. O. Vartia. Ideal Log-Change Index Numbers. *The Scandinavian Journal of Statistics*, 3: 121-126, 1976.

B. Vidakovic. *Statistical Modelling by Wavelets*. Wiley, New York, 1999.