Commodity Prices in the Cape Colony, 1889-1914

Laurie Binge[1]
Willem H. Boshoff[2]
Stellenbosch University
This draft: January 10, 2017

Commodity prices in the Cape Colony...

JEL Classification: C
Keywords: Cape Colony, Commodity Price Index, Repeat Sales

1 Introduction

The turn of the 20th century was a period of huge structural change in South Africa. The discovery of gold (1886) on the Witwatersrand fundamentally altered the structure of the economy from essentially pastoral to mining intensive. The Second South African War (1899-1902) broke out and fundamentally altered the region, and led to the formation of the Union of South Africa from the four colonies (1910). The seminal works of De Kock (1924), Schumann (1938) and De Kiewiet (1941) discuss the economic and social history of the period.

More recently, there has been what Fourie (2016) calls a Data Revolution in African Economic History. Over the past two decades the increase in access to online resources, data-processing software and computing power has enabled scholars to capture and analyse historical statistics on a much larger scale than before. Colonial-era archival records are being digitised and transcribed on a much larger scale. One example is the Colonial Blue Books, a vast collection of reports, containing records of the civil establishment, revenue and expenditure, imports and exports, and a range of other statistics. These data sources can now be used to study a range of issues, including population size, wages, incomes, education, fiscal systems, and transport networks of African societies. A range of studies have investigated this fascinating period using newly digitised historical data (see for instance, Boshoff and Fourie (2015), Greyling and Verhoef (2015) and Zwart (2011)).

This paper attempts to make a specific contribution to the quantitative history of the Cape Colony over this period. The first aim is to create a high frequency (i.e. monthly) wholesale price index for agricultural commodities for the Cape Colony for the period 1889-1914. This is a particularly interesting period of South African history, between the discovery of gold in the Witwatersrand (1886) and the outbreak of WWI (1914). Higher-frequency price indices are essential in studying business cycles, financial crises and market behaviour in general.

[1]PhD candidate at the Department of Economics at the University of Stellenbosch.
[2]Associate Professor at the Department of Economics at the University of Stellenbosch.

The aim is to provide some insight into the price history of selected episodes in South African economic and financial history.

The index is based on two new datasets that have been digitised and transcribed. The first is a dataset of monthly market prices of agricultural products in various towns reported in the Agricultural Journals of the Department of Agriculture. The second is a dataset of annual market prices of agricultural products in various town reported in the Cape Colony Blue Books. There are two challenges in creating a monthly wholesale price index for the period. The first is that the monthly data is incomplete, both in terms of the coverage of products and towns, and in some cases the Journal volumes that are just missing altogether. The second challenge is to incorporate annual prices in order to supplement the monthly price information.

Klovland (2014b) showed that the repeat sales method, which is more typically used in creating indices for infrequently traded goods such as real estate, may be used to deal with incomplete data in the context of constructing a relatively high-frequency historical price index for commodities. To supplement the indices Klovland (2014a) suggested temporally disaggregating (interpolating) annual data series to incorporate the annual series into the monthly series.

This paper will apply these techniques to estimate a historical commodity price index for the Cape Colony for the period 1889-1914. Wholesale indices are computed for each product. The product indices are then aggregated to form indices of product categories, using census data, as well as an overall wholesale commodity price index. The estimated monthly wholesale price indices are then compared to both existing price indices in South Africa and similar wholesales price indices in other countries to gauge if they provide reasonable results. The prices may be useful for further analysis of issues in macroeconomic and financial history, and convergence or market integration.

The prices are also used to analyse market integration over this period, where the mineral discoveries led to the substantial development of transport infrastructure and especially the railways. To this end the two datasets may be integrated in a slightly different way, using the same techniques. Two price series for a specific product town combination, e.g. wheat in Cape Town, can be combined with the repeat sales method to form a more continuous price index. Market integration is then investigated by looking at correlations in the prices of specific products for different regions and towns. The results show that the development of the railways led to increased market integration in the Cape Colony after the completion of the railway lines.

2 The Economic History of the Cape Colony

The seminal works of De Kock (1924), Schumann (1938) and De Kiewiet (1941) discuss the economic and social history of the South African colonies in the late 19th and early 20th centuries. These studies have similar narratives of the economic development of South Africa around that period. More recently, Greyling and Verhoef (2015) calculated GDP estimates for the Cape Colony, in order to facilitate the analysis of long-term growth and development in the Cape. This section provides a brief description of the economic history of the period (i.e. GDP growth, inflation, imports and exports), by synthesising these narratives and relating them to the quantitative evidence from the GDP estimates.

Figure 1 illustrates the real GDP estimates for the Cape Colony (1956-1909) reported in Greyling and Verhoef (2015), with recessionary periods shaded. They derived the business cycle by decomposing the GDP series into its cyclical and trend components, using the Baxter-King band-pass filter. An upswing phase was defined as a period of actual GDP higher than trend GDP and a recession as a period of actual GDP lower than trend GDP. The turning points should be taken as indicative, as there are many ways to date business cycle turning points, which would lead to slightly different turning point dates.

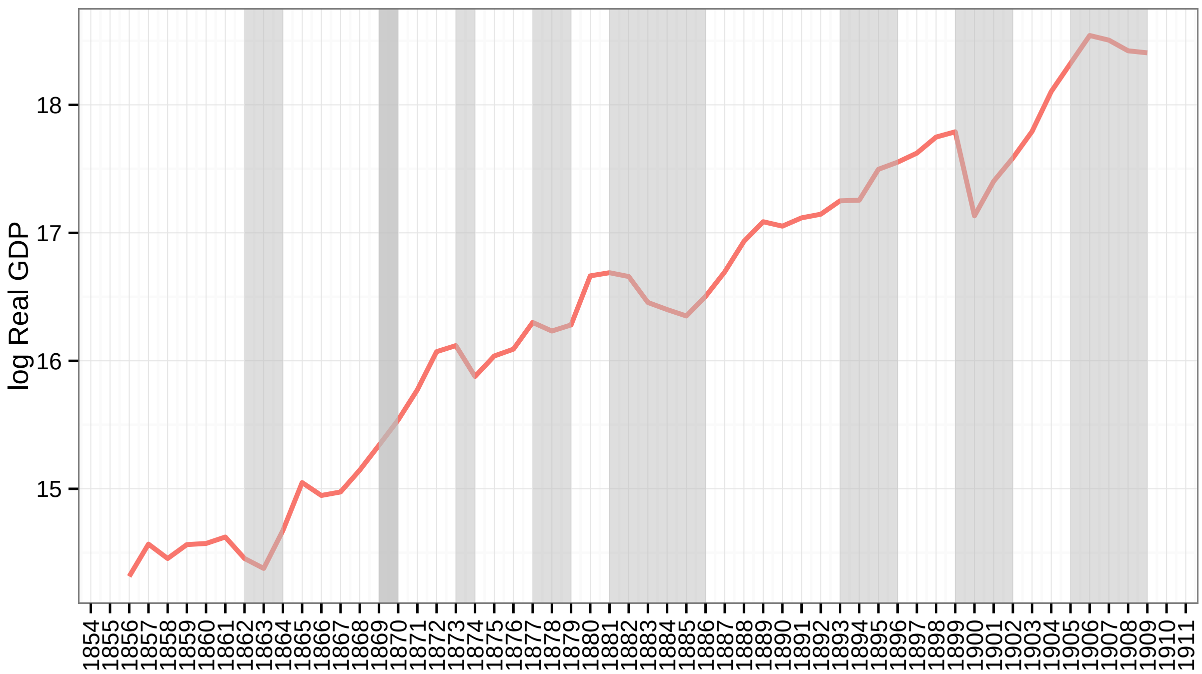

Figure 1: Real GDP in the Cape Colony (1956-1909)

The focus here is on the last two decades of the period, between the discovery of gold in 1886 and the formation of the Union in 1910. Taken as a whole the Cape Colony experienced relatively rapid expansion over this so-called "gold-mining" period, during which the structure of the economy transformed from an agricultural economy to an agricultural-mineral economy (Schumann, 1938). According to the GDP estimates, the Cape Colony experienced three upswing phases and three downswing phases between 1886 and 1909, with the most severe recession during the Second South African War (1899-1902).

According to the GDP estimates, the discovery of gold was followed by an upswing phase, lasting roughly from 1886 to 1892. According to the historical sources (e.g. Schumann (1938) and De Kock (1924)), the upswing phase was followed by a depression as the result of a speculative financial and banking crisis in 1889-1890. The banking crisis was due to an overextension of credit, and speculation in gold shares and property. Most of the local or district banks in the Cape got into financial difficulties and were absorbed by the larger banking institutions (De Kock, 1924). The price of gold shares, land and prospecting companies plummeted, while railway traffic, government revenue and imports declined marginally (Greyling and Verhoef, 2015). The crisis was intensified by the Baring crisis of 1890 in England. However, Schumann (1938) reports that the effect on the banks of the Cape was much more severe than on economic conditions in general. Agriculture in the Cape Colony hardly felt the depression, as the harvest prospects in 1890 were excellent. This is reflected in the flat real GDP figures from 1890 to 1893, when a peak was reached.

The recession lasted from 1893 to 1896 according to the GDP estimates. Political and natural factors contributed to the relatively stagnant period (Gilbert, 1933). The Jameson Raid at the end of 1895 had an unsettling effect on business conditions. Tensions between the Boer Republics and Britain led to the closure of the drifts in September and October 1895, disrupting road transport of goods into the interior or outward to the coastal ports. A severe drought in 1896 and the rinderpest cattle disease caused widespread livestock losses (De Kock, 1924; Schumann, 1938).

According to the turning point dates, there was a mild recovery between 1896 and 1899, although GDP was still relatively stagnant. As the country was slowly emerging from this crisis, the outbreak of the Second South African War (1899-1902) caused a severe recession.

The War had a devastating effect on the economies of the Boer Republics. Agricultural prospects were severely hampered following the "scorched earth" policy of the British forces. This policy caused destruction of farm buildings, crops and livestock in the Transvaal and the Orange Free State, and to a lesser extent in parts of the Cape Colony and Natal. Gold mining in the Transvaal came to an almost complete standstill until 1902, with the value of gold production declining from £16m in 1898 to £1m in 1901 (Schumann, 1938).

The huge increase in expenditure by British Government in connection with military operations created prosperity in parts of the Cape Colony and Natal, especially amongst the farming and trading communities. The war cost Great Britain around £250 million, of which a large proportion was spent in the Cape Colony and Natal to purchase agricultural and pastoral produce for the troops. Within a few months of the outbreak the War, the number of British soldiers operating in South Africa increased to 130,000, and subsequently to around 250,000 (nearly 25% of the white population of the four territories combined) (Schumann, 1938).

The War demand for goods and services stimulated increases in price and production (De Kock, 1924). Due to the increased local demand for all kinds of produce, exports from Cape and Natal ports also declined considerably, from almost £25m in 1898 to £7.5 in 1900. Figure 2 illustrates the decrease in exports, as well as the marked increase in imports from 1899 to 1903, due to in large part to the increased imports of military supplies for the British troops (Gilbert, 1933). British military expenditure on around 250,000 soldiers therefore provided an injection for consumption and production in the Cape Colony and Natal (Greyling and Verhoef, 2015). According to the Cape Colony GDP estimates, the trough was already reached in 1900, which was followed by quite a swift recovery.

The end of the War in 1902 was followed by an upswing phase. According to Schumann

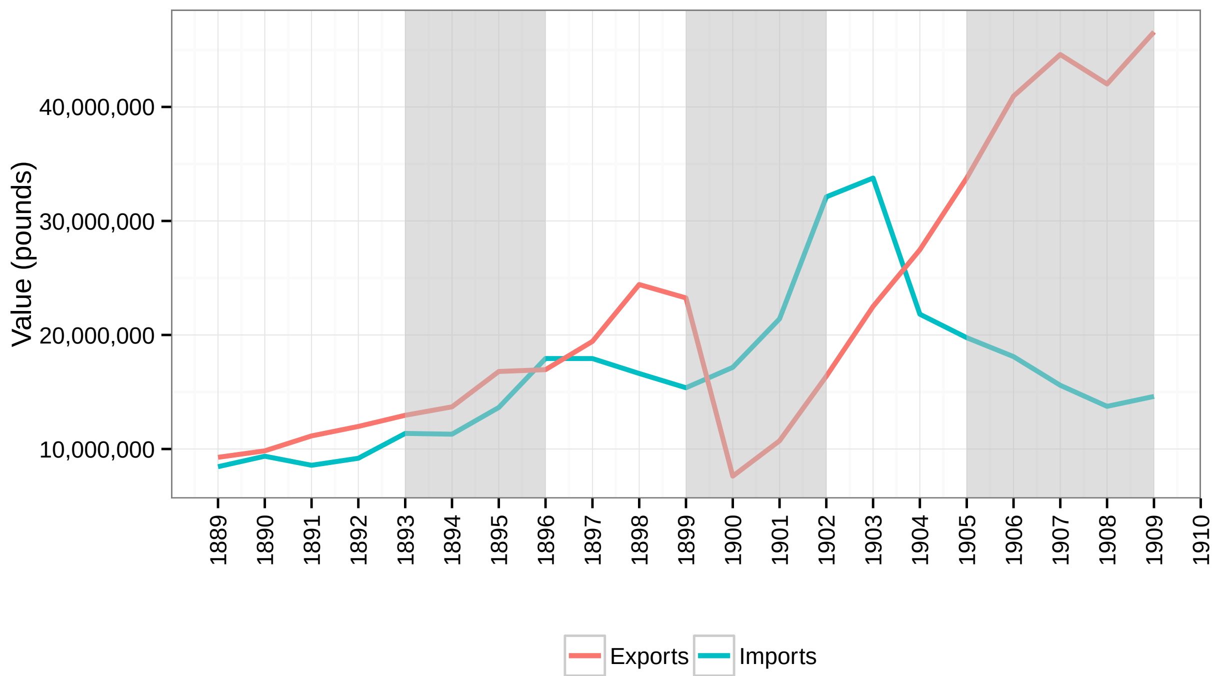

Figure 2: Total Imports and Exports in the Cape Colony (1986-1909)

(1938), the post-war boom was mainly the result of a general feeling of optimism, both in England and in South Africa, after the conclusion of the war. This optimism resulted in the extension of bank credit and extensive speculation in property, as well as an increase in the importation of capital and goods to aid reconstruction. The reconstruction schemes and the Development Loan of £3 million granted by the British Government, for instance, stimulated economic activity and a speculative environment (De Kock, 1924). There was also large immigration, presumably in search of mineral wealth. For example, 71,081 people landed at Cape ports in 1903, but by 1908 the number had decreased 27,192. Exports also increased significantly over the period (Gilbert, 1933).

However, the post-war boom was short-lived. According to the GDP estimates, recession set in by 1905, although historical sources put the date slightly earlier in 1903. According to Schumann (1938), the depression was caused by a cyclical reaction to the large economic disequilibrium. Production was expanded after the war, based on the expected expansion of the gold and other industries. But when the expectations were not fully realised, a reaction set in and a long period of depression followed (De Kock, 1924). The inflow of purchasing power due to the war ended and large stocks of military goods were sold. The depression may be explained in part by the complete destruction of the interior, despite the relatively early return to gold production (Greyling and Verhoef, 2015). The progress of the gold industry was hampered by the lack of an adequate supply of unskilled labour after the War (De Kock, 1924). The financial panic in Europe and America in 1907 and its effects on South African trade, as well as the drought of 1908, may also have deepened the depression (De Kock, 1924).

This was a tumultuous period of South African history. The biggest event seems to have been the South African War. The new data provides a way of anchoring and illustrating the narrative. The following section introduces the two datasets.

3 The Data

The first aim of the paper is to estimate a monthly historical commodity price index for the Cape between 1889 and 1914. The choice of period is constrained by the availability of high-frequency data. The wholesale commodity price indices presented below are based on two new datasets that have been digitised and transcribed.[3]

The first is a dataset of monthly "*Current Market Rates (Wholesale) of Agricultural Produce*", as telegraphed by the Civil Commissioners and reported in the Agricultural Journals of the Department of Agriculture of the Cape Colony (available in the Elsenburg library). The Journals are available from October 1889 to August 1914, which dictates the period that the indices will cover. Monthly data are available for 24 different commodities (e.g. wheat, eggs and tobacco)[4] in 20 towns across the Cape Colony (e.g. Cape Town, Kimberley and Port Elizabeth).[5]

Figure 3 illustrates the number of monthly observations by commodity. Clearly the sample suffers from missing data points. Not only do a number of commodities and towns disappear from the recorded sample, but certain volumes of the Agricultural Journals are missing altogether (i.e. Jan97-Jun97, Sep97, Jul98-Dec98, Jan00-Jun00, Jul02-Aug02, Apr05-May05, Aug05, Jul06-Jun07, Nov08, Jul09-Dec09, Apr10-May10, and Jan11-Feb12). The increase in observations around the turn of the century is due to the fact that weekly prices were recorded in the reports over that period.

Figure 4 illustrates the monthly prices for one of the commodities, wheat, by Cape Colony town. Wheat prices varied widely over the period, especially around the turn of the 20th century. The challenge is to combine these time series in a consistent way in order to construct a coherent index of monthly wheat prices for the Cape Colony.

The second dataset includes average annual market prices reported in the Colonial Blue Books of the Cape Colony (available from the British Online Archives). In each annual volume, data are presented for the average prices of many commodities from 1889 to 1907 for various towns in the Cape Colony. The annual market prices in Colonial Blue Books were collected by the

[3]All prices were reported in pounds sterling (£), shillings (s) and pennies (d) and converted to pennies. There were twenty (20) shillings (s) per pound (£). The shilling (s) was subdivided into twelve (12) pennies (d).

[4]The 24 products are wheat, wheat flour, boer meal, mealies, mealie meal, barley, oats, oathay, lucerne hay, potatoes, tobacco (boer roll), beef, mutton, fresh butter, eggs, cattle (slaughter), sheep (slaughter), pigs (slaughter), bread, oranges, saddle-horses, transport oxen, milch cows, woolled sheep.

[5]The 20 Cape Colony towns are: Aliwal North, Beaufort West, Burghersdorp, Cape Town, Clanwilliam, Cradock, Dordrecht, East London, Graaff-Reinet, Graham's Town, Kimberley, King William's Town, Malmesbury, Mossel Bay, Port Alfred, Port Elizabeth, Queen's Town, Tarkastad, Vryburg and Worcester. The data also includes a few prices for towns in the other territories (e.g. Johannesburg) in the final few months of the sample. Unfortunately, these are too few to include in the indices.

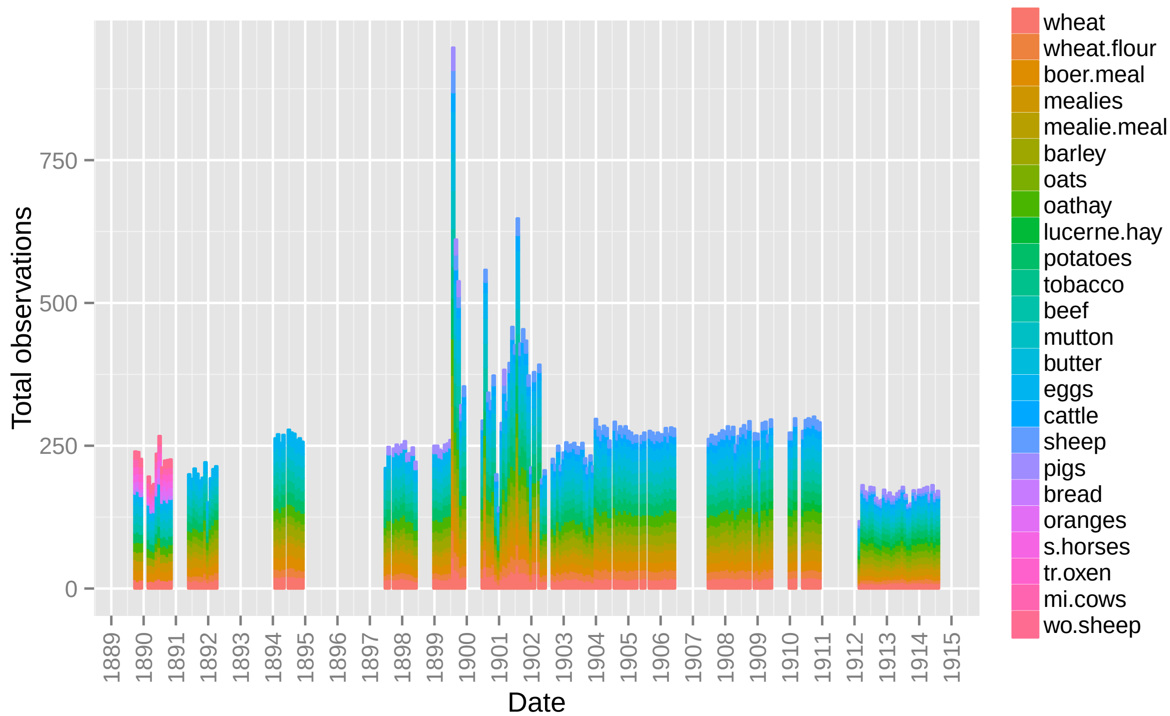

Figure 3: Total the number of monthly observations by commodity

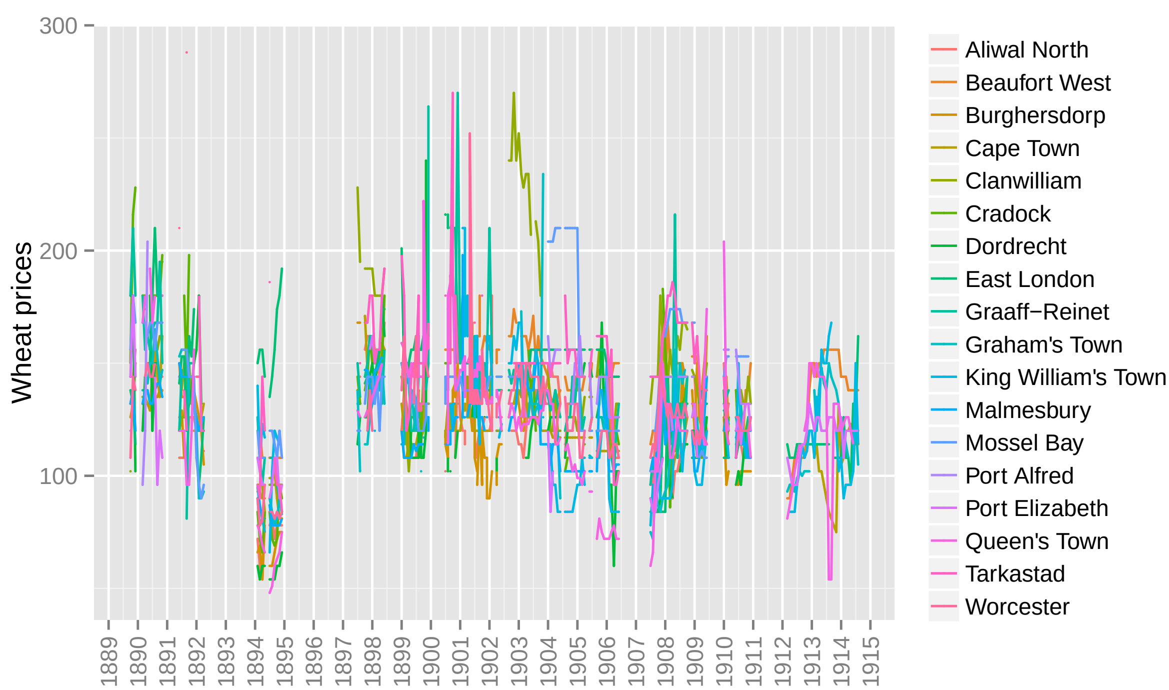

Figure 4: Monthly wheat prices by town - Agricultural Journals

colonial administrators in South Africa and sent to the Colonial Office in London (Zwart, 2011). This information on "Average Market Prices of Agricultural Produce,"[6] "Provisions,"[7] and "Stock and Animal Productions"[8] were used to create a database that includes prices of 50 agricultural products in 48 towns[9]. In a few cases there are more than one series for a specific commodity, such as three different series for wine (ordinary quality, better quality and just wine without a description) and two series for beer (English and Colonial).

Again, a number of commodities and town have missing observations in the sample. The annual data is incomplete in terms of the coverage of products and towns. Figure 5 illustrates the annual prices for wheat by town. Again, wheat prices seem to have varied widely between towns in the Cape Colony.

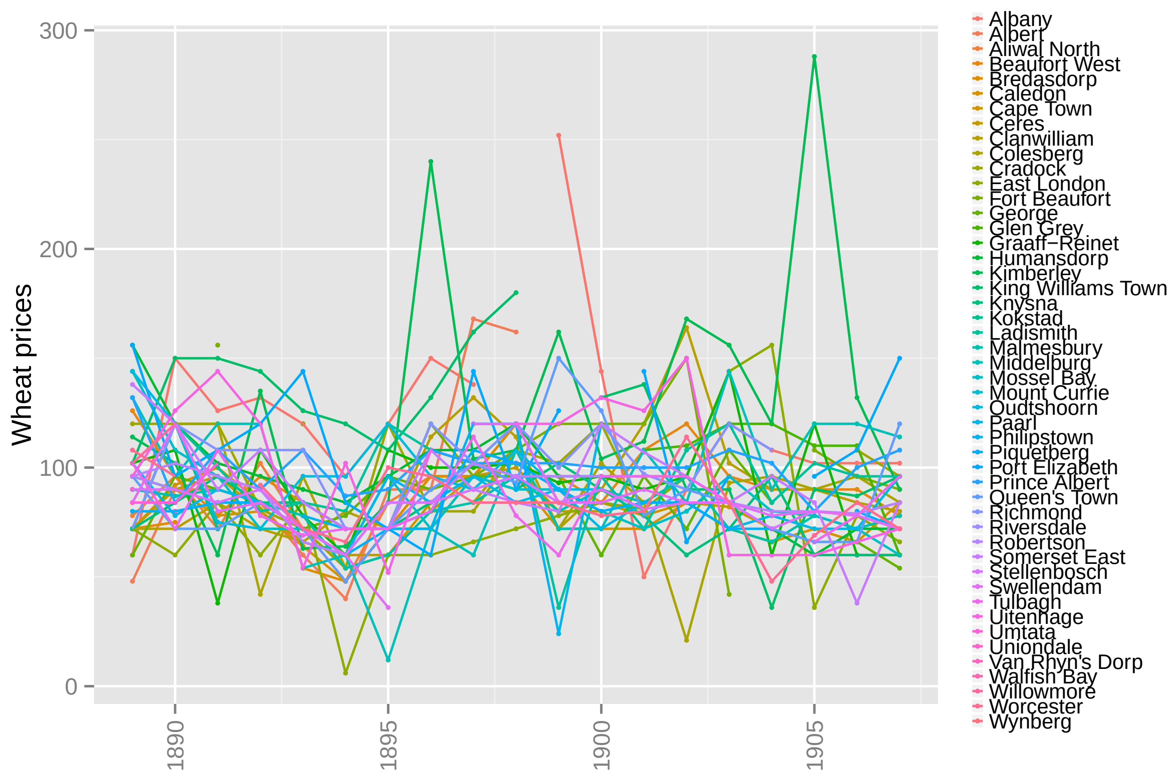

Figure 5: Annual wheat prices by town - Blue Books

[6]Wheat, barley, rye, oats, mealies, peas & beans, potatoes, pumpkins, aloes, argol, wine and brandy.

[7]Oatmeal, flour, bread, mutton, beef, pork, bacon, butter, cheese, tea, coffee, sugar, rice, tobacco, dried fruit, salt, wine, brandy, beer, milk, candles, and lamp oil.

[8]Saddle horse, draught mules, asses, draught oxen, milch cows, woolled sheep, cape sheep, swine, goats, fowls, ducks, washed wool, fat and tallow, soap, hides, sheep skins, and goat skins.

[9]The 48 Cape Colony towns are: Albany, Albert, Aliwal North, Beaufort West, Bredasdorp, Caledon, Cape Town, Ceres, Clanwilliam, Colesberg, Cradock, East London, Fort Beaufort, George, Glen Grey, Graaff-Reinet, Humansdorp, Kimberley, King Williams Town, Knysna, Ladismith, Malmesbury, Middelburg, Mossel Bay, Oudtshoorn, Paarl, Philipstown, Piquetberg, Port Elizabeth, Prince Albert, Queen's Town, Richmond, Riversdale, Robertson, Somerset East, Stellenbosch, Swellendam, Tulbagh, Uitenhage, Uniondale, Willowmore, Worcester, Van Rhyn's Dorp, Wynberg, Walfish Bay, Mount Currie, Kokstad, Umtata.

In some cases the commodity and town series in the two datasets overlap, although they are almost always reported in different units. This means that the trends can be compared, but the levels are different. Figure 6 illustrates the case of wheat prices in Cape Town. In the monthly dataset wheat prices are reported in pounds (lbs), whereas in the annual data the wheat prices are reported per bushel. The average prices reported in the two datasets seem to capture a similar trend over time, as one would expect.

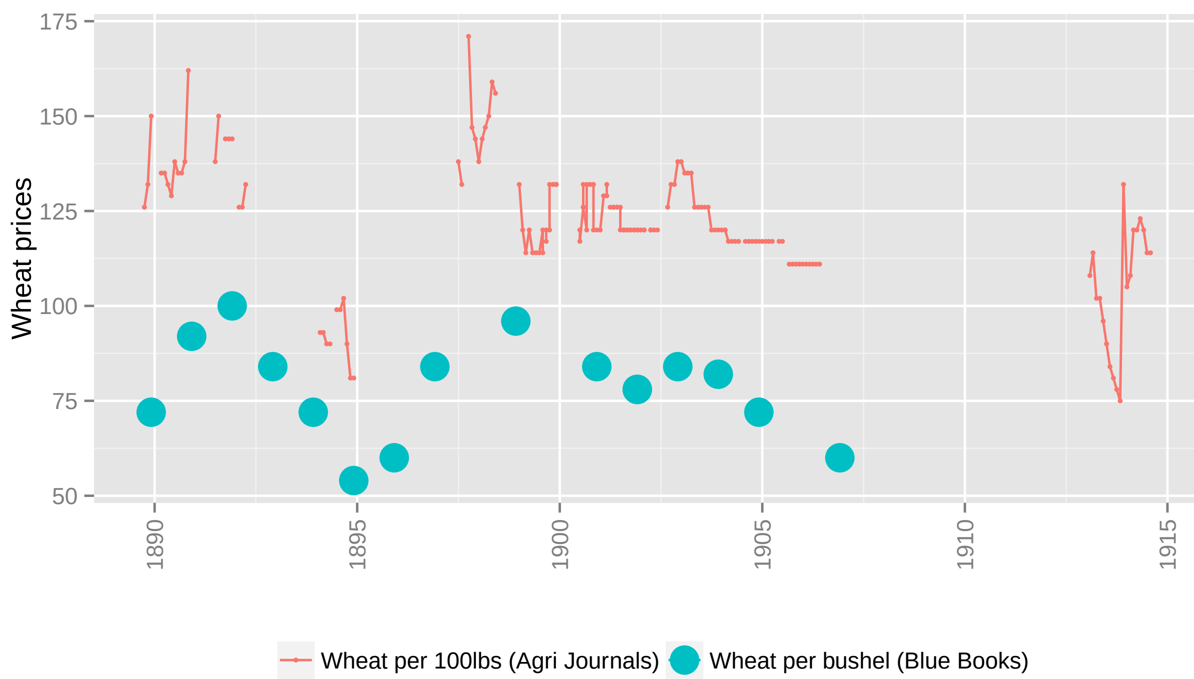

Figure 6: Wheat prices in Cape Town

The challenge then is to combine and aggregate these two different sets of prices in a consistent way in order to create coherent price indices of agricultural produce for the Cape Colony. This is complicated by the fact that there are so many missing values in both panels. The following section presents the methodology suggested to tackle these challenges.

4 Methodology

4.1 Price indices

The aim is to use these two datasets of market prices to construct a monthly historical commodity price index for South Africa for the period 1889 to 1914. Higher-frequency price indices are useful in studying business cycles, crises and market behaviour in general. As far as the author is aware no such commodity price index exists for the Cape Colony over that period.

The first generation of historical price indices, such as Jevons (1865) and Giffen (1879), were focused on the influence of the supply of precious metals (typically gold) on the variation in the general price level. Gilbert (1933)'s analysis of the economic effects of the gold discoveries upon South Africa follows this tradition. In this study he computed an annual weighted aggregative index of the prices of 13 foods for 1883-1907. Another strand in the literature, including Persons and Coyle (1921) and Silberling (1923), focused on the measurement of business cycles or "*business conditions during alternating periods of prosperity and depression.*" Schumann (1938)'s seminal work on business cycles in South Africa follows this tradition. He constructed a quarterly wholesale price index for this analysis from 1910-1936.

The brief review of the data sources showed that the price series were extracted from publications that were incomplete and in some cases just missing. This is a common occurrence, as many historical price series must be extracted from publications which are incomplete and in some cases no longer accessible. Even if sources are available on a continual basis there will inevitably be cases when a particular data series has been discontinued or contains substantial gaps. The supply of agricultural commodities may vary according to seasonal factors and occasional crop failures, creating gaps in the price series. Coverage may deteriorate during periods of conflict or war. Quality descriptions and publication practices may also change over time. These problems make it challenging to combine price data to form continuous time series (Klovland, 2014b).

The traditional way of dealing with the problem of gaps in historical data is to splice the time series at a point in time when there is overlapping information (see for instance Solar and Klovland (2011)). This method requires great care and involves substantial work. If there are many time series and various gaps in the data, as is often the case with higher frequency historical data, this procedure may become very difficult to implement consistently (Klovland, 2014b).

Klovland (2014b) recommended that the repeat sales method could be useful whenever underlying series were characterised by incomplete observations. The repeat sales method was specifically developed for a market where the price of each object is quoted infrequently and at irregular intervals, which is typical in the real estate market. A similar, albeit less extreme, situation is typical of historical price data, where the gaps between the observed prices are usually shorter. The problems encountered in aggregating the individual time series to an overall index are in principle the same. One can think of each commodity price series as a specific asset (e.g. an artwork). The repeat sales approach compares that same commodity series over time, with a sale occurring whenever a price data point is observed.

4.2 Repeat Sales Method

The repeat sales method was initially proposed by Bailey, Muth and Nourse (1963a) to calculate house price changes. They saw their procedure as a generalisation of the chained matched model methodology applied previously in the construction of real estate price indices. The method was subsequently extended by Case and Shiller (1987), among others, and is currently used to produce the S&P/Case-Shiller Home Price Indices in the US.

The repeat sales method tracks the sale of the same item over time. It aggregates sales pairs and estimates the average return on the set of commodities in each period (Kräussl and Lee, 2010). The index numbers are derived directly from a least squares regression involving a set of dummy variables and requiring only information on the rate of change of the individual price series and the dates at which the data are observed (Klovland, 2014b).

Following Wang and Zorn (1997), the model underlying Bailey, Muth and Nourse (1963b)'s method may be derived in the following way. Where t_1 and t_2 are the times of the first and second transactions, a price relative $\frac{P_{i,t_2}}{P_{i,t_1}}$ can be modelled as:

$$\frac{P_{i,t_2}}{P_{i,t_1}} = \frac{I_{t_2}}{I_{t_1}} \times u_{i,t},$$

where P_{it} is the price of commodity i at time t; I_t is the true but unknown index for period t; and $u_{i,t}$ is the idiosyncratic error term. Taking logs,

$$\ln \frac{P_{i,t_2}}{P_{i,t_1}} = -\ln(I_{t_1}) + \ln(I_{t_2}) + \ln(u_{i,t})$$

Using vector notation, this relationship can be expressed as $y = X\beta + \epsilon$, where y is the (known) vector of logged price relatives; β is a T-dimensional column vector of unknown logarithms of the index numbers to be estimated, such that the t-th component of the (to this point unknown) β vector is $\beta_t = \ln(I_t)$; X is a matrix of $(n \times T)$ dimensions such that the t-th component of each row is -1 if $t = t_1$, +1 if $t = t_2$, and 0 otherwise; and ϵ is the vector of $\log(u_{i,t})$ values. The repeat sales model may also be derived as the first differences of a hedonic model.

The standard specification of the repeat sales usually takes the following form:

$$\ln \frac{P_{it}}{P_{is}} = \sum_{t=1}^{T} \beta_t D_{it} + u_{it}$$

where P_{it} is the price of a particular commodity i (e.g. wheat in Cape Town) in time t; P_{is} is the price of exactly the same commodity i at time s; β_t is the parameter to be estimated for time t; D_{it} represents a time dummy equal to 1 in period t when the resale occurs, -1 in period s when the previous sale occurs, and 0 otherwise; and u_{it} is a white noise residual.

Thus, in the standard repeat sales model the dependent variable is regressed on a set of dummy variables corresponding to time periods. The coefficients are estimated only on the basis of changes in item prices over time. The index numbers are derived from the dummy variable coefficients, requiring information only on the rate of change of the individual price series and the dates at which the data are observed. The price index is simply the antilog of the series of estimated coefficients: $\hat{\beta}_1, ..., \hat{\beta}_T$.

At each time t there exists a distribution of growth rates in the population. The repeat sales estimator is a measure of the central tendency of this distribution, in the form of the geometric mean of the growth rates of the items that sold more than once.[10] Bailey, Muth and Nourse (1963a) showed that the β estimates (i.e. the mean logged price indices) consist of the period-by-period weighted averages of the logged price relatives, with weights proportional to their sample sizes. In other words, the regression solutions are complicated weighted averages of the average logged price relatives. Wang and Zorn (1997) showed this clearly for a relatively simple and intuitively appealing example with two periods. For example, the first index value in t_1 is the weighted average of the two quantities: the average logged appreciation of items observed in period t_0 and t_1; and the average appreciation from the t_0 to t_2 minus the average appreciation from t_1 to t_2. The weights given to the two estimators are proportional to their sample sizes.

In this case, the repeat sales method has the great advantage of being able to handle gaps in the data series of any length. It utilises all of the information in the dataset, compared to the cumbersome manual splicing of time series of individual prices of the traditional approach. As Wang and Zorn (1997) noted, if the number of observations does not vary across time periods (i.e. if there are no gaps in the time series), the repeat sales estimator simplifies to an ordinary chain index.

One would like to observe prices for a static commodity description over time (e.g. English beer per bottle). However, new qualities are often introduced to the market (e.g. beer per gallon) and old ones disappear. For some commodities the sources of supply may vary by season. The repeat sales approach can also handle this case by tracking prices for the same product description over time.

The typical disadvantages of the repeat sales method are not real concerns in this case, namely that single-sale data is discarded and that there is sample selection bias in the types of commodities that are sold more than once (Hansen, 2009). In this cases all of the commodities were sold more than once and can be included in the index estimates. It is also possible to attach less weight to observations calculated from rates of change over long periods of time than on changes from adjacent periods. For instance, changes in product specifications or other characteristics are more likely to have changed if there are large time intervals between the observations. Case and Shiller (1987) suggested a weighted three-step least square procedure to deal with this problem However, this is likely to be less of a problem in this sample because the sales appear so frequently, compared to real estate prices which often only resell after decades.

4.2.1 Repeat Sales Example

A simple example from the data may be used to illustrate the procedure. Table 1 reports the wheat prices per 100lbs in 3 towns over a period of 10 months. The purpose is to calculate

[10] Such an index will track the geometric mean, rather than the arithmetic mean, of prices over time, because of the log transformation prior to estimation. If it is assumed that the regression residuals are normally distributed in each period, a correction can be made by defining corrected index values as: $I_t = \exp\left[\gamma_t + 1/2(\sigma_t^2 - \sigma_0^2)\right] * 100$, where σ_t^2 is the estimated variance of the residuals in period t (Renneboog & Spaenjers 2012). In practice this adjustment is often negligible (Hansen 2009), which is also the case in this sample.

a price index for this period, given the missing observations (NAs) in the data. The main challenge is to put the data together in a consistent way. For period 3 it would be natural to combine the price increases from Beaufort West and Cape Town. For the other periods there are a number of ways to splice the time series. In practice, with many time series, such procedures would be very difficult and time consuming to handle consistently.

Table 1: Repeat sales example with wheat prices

Date	Period	Beaufort.West	Cape.Town	Worcester	Index
Jun 1891	1	150		210	100.00
Jul 1891	2	135	138		88.83
Aug 1891	3	135	150		92.01
Sep 1891	4			288	138.94
Oct 1891	5		144		86.37
Nov 1891	6		144		84.45
Dec 1891	7	120	144		82.57
Jan 1892	8			144	70.38
Feb 1892	9		126	144	71.31
Mar 1892	10		126		71.31

The repeat sales method uses a simple least squares regression on time dummies to produce the estimated index values. Repeat sales pairs are formed from the price relatives of the same commodity (in this case wheat) for each town. For example, for Beaufort West, 3 sales pairs can be formed (e.g. 135/150). The sample consists of 12 sales pairs and 10 time dummies are created for the 10 time periods, reported in Table 2. The first sales pair corresponds to the first two observations for Beaufort West, with a price decrease of 10.5% between period 1 and 2 and time dummies of -1 and 1 for the first two periods. The other rows are derived in the same manner. Running a least squares regression on this data set produces a set of coefficient estimates for the 10 dummy variables. Taking the antilog to these estimates produces the index in the table above. In this case the additional prices from Worcester help to make the index continuous, specifically by adding price observations in the 4th and 8th periods, where there were only missing observations. Thus, including more price series will lead to fewer gaps in the index.

This approach does treat all observations as equal, which means that variables observed more frequently than other ones will exert a stronger influence on the index, simply because there will be more observations in the data set originating from this variable. For example, if wheat prices from Cape Town are recorded and included in the index calculations more frequently than Worcester prices, the Cape Town prices will have a larger implicit weighting in the index. This might be a sensible approach, as regularly quoted prices are often the ones most frequently traded (Klovland, 2014b). In other words, the prices of the same commodity in different towns are treated equally and towns with more observations have an implicit higher weighting. Even if the prices are at different levels due to factors such as transport costs, for the purposes of the index the trends (or growth rates) are being compared over time.

It is possible that the smaller towns have more observations in some periods that do not properly reflect price movements in the larger towns, which might then bias the index to some degree. There might be an idiosyncratic spike in Worcester wheat prices in a period where

Table 2: Rgression input of repeat sales example with wheat prices

ln(Pt/Ps)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
-0.105	-1	1	0	0	0	0	0	0	0	0
0.000	0	-1	1	0	0	0	0	0	0	0
-0.117	0	0	-1	0	0	0	1	0	0	0
0.083	0	-1	1	0	0	0	0	0	0	0
-0.040	0	0	-1	0	1	0	0	0	0	0
0.000	0	0	0	0	-1	1	0	0	0	0
0.000	0	0	0	0	0	-1	1	0	0	0
-0.133	0	0	0	0	0	0	-1	0	1	0
0.000	0	0	0	0	0	0	0	0	-1	1
0.316	-1	0	0	1	0	0	0	0	0	0
-0.693	0	0	0	-1	0	0	0	1	0	0
0.000	0	0	0	0	0	0	0	-1	1	0

only Worcester prices were reported. For example, in period 4 in the example above, the index spikes largely because of the Worcester price. However, idiosyncratic price movements in specific towns are less likely to bias the overall indices because the dataset includes a large number of towns from two separate datasets.

The repeat sales method therefore provides a consistent way to aggregate the data from the different towns for a specific commodity. This method also generates an index with substantially fewer gaps than there are in the individual series.

4.2.2 Creating indices for specific commodities

Klovland (2014b) suggested that the simple unweighted version of the repeat sales model is most applicable at the lowest level of aggregation, i.e. for specific commodities such as wheat, when explicit weighting of different price observations is less crucial. The towns are weighted implicitly according to coverage, i.e. the number of observations in each dataset.

Following this suggestion, the repeat sales method is used to aggregate the different time series for each commodity from the all the towns in the Cape Colony. This is the method described in the example above. This method may be applied to both the monthly dataset and the annual dataset separately, to obtain two more complete indices for each commodity, one monthly and one annual. In other words, the repeat sales method can similarly be applied to the annual price series, in order to form more complete annual price indices for each commodity.

In this way the annual prices from the Blue Books can be used to supplement the results from the monthly dataset. This is useful because the monthly commodity indices still contain substantial gaps, even after the repeat sales method has been implemented, where no data is available. It is also useful because the monthly dataset omits a number of important products. Wool is one example of an important export product in the Cape Colony, which is not present in the monthly dataset.

The challenge then is to incorporate the lower frequency (i.e. annual) index into a higher frequency (i.e. monthly) index. The Blue Books data reports the annual market prices as average prices for November of each year. The simplest way to incorporate the annual prices is to set them as the November prices for each year, and then simply to interpolate between the data points to form a monthly index.[11]

The two monthly indices for each commodity may then be combined again with the repeat sales method in order to obtain one reasonable continuous index for each commodity which covers all the towns in the Cape Colony. The indices from the two data sources therefore implicitly get equal weighting in this step. And because the annual data has already been interpolated, the monthly series is also interpolated at this stage. This involves a significantly smaller amount of interpolation than would have been necessary if each individual series were interpolated from the start.

Figure 7 illustrates an application for wheat prices in the Cape Colony. In the first step the monthly wheat prices for 20 towns in the Cape Colony are combined to form a monthly wheat index, which is then interpolated to form a continuous series. The annual wheat prices for the 48 towns are also combined to form an annual wheat price index, which is then interpolated to form a monthly wheat price index. These two indices are combined using the repeat sales method to form the overall monthly wheat price index for the Cape Colony. Thus, by combining the data a time series with more complete coverage can be obtained. The interpolation that occurs at this stage is much less severe than would be the case if all of the individual price series were interpolated from the start. However, there are still gaps in the index in the periods where there was no price information available.

4.3 Temporal Disaggregation

Wimpie, ek het 'n lang stuk gehad oor temporal distribution. Maar aangesien dit nie so lekker werk nie, wonder ek of mens dit moet uithaal. Hier is 'n verkorte weergawe, sodat jy kan sien waarvan ek praat.

The process of deriving high-frequency data from low-frequency data and, if available, related high-frequency data is called temporal disaggregation. There are two facets to the temporal disaggregation problem: temporal distribution and interpolation. Both concern the estimation of intra-period values of a variable whose actual values are observed only once per period. Temporal distribution, which arises with flow variables (or the average of stock variables), is the estimation of several values, the sum or average of which equals the observed value over the longer term. Interpolation refers to the estimation of a stock variable whose actual values are observed less frequently (Litterman, 1983).

All disaggregation methods ensure that either the sum, the average, the first or the last value of the resulting high-frequency series is consistent with the low-frequency series. Temporal

[11] There is a technical distinction between temporal distribution and interpolation. On the one hand, the distribution problem appears when the observed values of a flow low-frequency series of length N must be distributed among kN values (where k is the number of sub-periods in which each period of low-frequency is divided. On the other hand, the interpolation problem consists in generating a high-frequency series with the values of the new series being the same as the ones of the low-frequency series for those temporal moments where the latter is observed (Chow 1971).

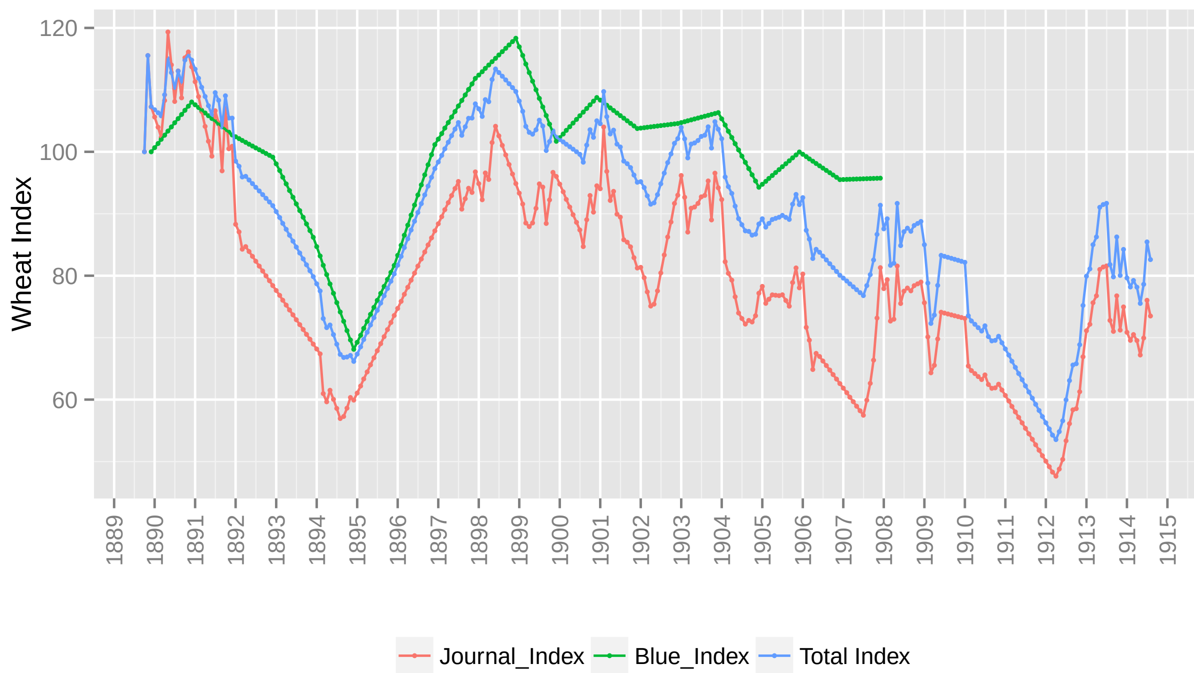

Figure 7: Wheat indices for the Cape Colony

disaggregation can be performed with or without one or more high-frequency indicator series. For example, quarterly exports could help disaggregating annual sales (Sax and Steiner, 2013). The goal is to obtain sub-annual estimates that preserve as much as possible the short-term movements in the indicator series under the restriction provided by the annual data which exhibit long term movements of the series (Chen, 2007).

A wide variety of procedures have been proposed in the statistical and economic literature to solve the problem of transforming a low-frequency series into a high-frequency one, including Denton, Denton-Cholette, Chow-Lin and Litterman. Klovland (2014b) uses the procedure suggested by Litterman (1983) as smoothing algorithm for annual prices. It uses related series (Klovland (2014b) used only a constant and a time trend) together with assumptions about the error term (e.g. 1st order serial correlation) to distribute the annual values over the twelve months of the year. In addition to smoothing the intra-year movements, this method ensures that the annual average of the estimated monthly data equals the true annual average.

In most cases the Blue Books report the prices to be the November prices. This means that it is more appropriate to set the annual prices to be the November prices for each year, and to interpolate them from there, rather than setting them equal to the annual average. This will create monthly series that are not as smoothed as if one would choose to distribute based on the annual averages. When annual prices are distributed based only on a constant, and the monthly prices are simply set equal to the annual values, which amounts to a form of simple interpolation between the data points. In effect the annual data are just interpolated between the annual data points, which are set to be the November prices.

The other alternative would be to distribute the annual data with reference to the monthly data for the same commodity, which would act as the indicator variable. However, there are a number of problems with this approach. The first problem is that the indicator variable only exists for a portion of the commodities, so it would only be applied for a small portion of the series. The second problem is that the indicator variable has to be complete. In many cases where the monthly data exists, it is only available for a fraction of the period. In all cases the monthly data would have to be interpolated for the entire period, and the annual data would be based on this indicator in the cases where it is possible. The most important problem is that in these cases the annual series virtually takes on the values of the monthly indicator series when those prices are available.

Figure 8 illustrates the procedure using actual monthly wheat prices for a certain period. The first two series are distributed based on a constant, constraining the last value or the average value to be equal to the annual value. In the third series the Litterman procedure was applied to the annual time series, distributing its values using the monthly wheat index as the indicator variable.

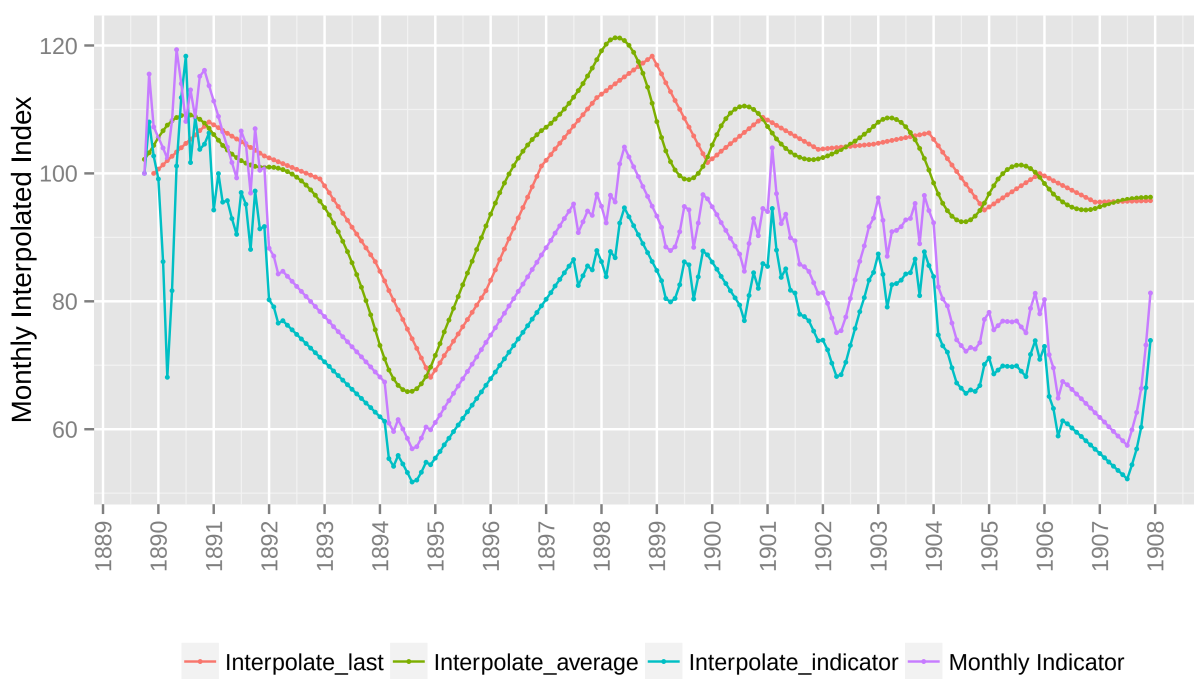

Figure 8: Temporal disaggregation of wheat prices

4.4 Aggregation of commodity indices

The construction of the indices involve three levels of aggregation. The first level involves aggregating the price information from the various towns and the two sources into indices for 43 commodities. This is done by using the repeat sales method described above. This method involves piecing together the individual data series on each commodity in an efficient way to form an index series by using all available price information. There is only an implicit weighting involved here, following from the fact that the descriptions with the greatest number of observations are the most influential ones in determining the coefficient estimates (Klovland, 2014b).

The next level is to aggregate the individual commodity price indices into indices for 7 larger commodity groups, which are weighted together in the conventional manner, e.g. with the Laspeyres price index. For instance, Klovland (2014b) aggregated his 110 commodities into 16 commodity groups (grain, meat, etc.) using chained Laspeyres indices, with weights based on output or trade values in 1835, 1870, 1890 and 1910.The different commodity indices for the Cape Colony are aggregated by applying weights based on the market shares of each good in domestic production. The value shares are used to obtain a weighted average index for all commodities in the Cape Colony. The final level then combines the indices for the commodity groups into an overall aggregate wholesale prices index.

This raises the question of what weights to use to aggregate the different commodities. The Agricultural Journals, particularly in the earlier volumes refer to these prices as "wholesale" rates. The term "wholesale" may be defined as prices charged for sales in large lots, usually at the first commercial transaction or in major trading centres (Klovland, 2014b). The first price indices were commonly referred to as wholesale price indices and typically comprised both domestically produced goods and imported goods. However, the principles of weighting were always applied consistently. More recently the focus has shifted towards producer price indices, which focus on prices obtained by domestic producers, and therefore domestic goods sold at home, and in some cases also exported goods (Klovland, 2014b).

In this case the focus is on domestic production shares and the produce returns reported in the 1904 census are used as the benchmark weights. The value shares were calculated as the quantity of production multiplied by the average market prices of each individual commodity. The difficulty is that the census did not cover all of the products and a number of important products would lack weights, e.g. wheat flour and beef. The weights for these products were based on imports values for specific products that were not reported in the census numbers. This weighting scheme make the index akin to Klovland (2014b)'s Wholesale Price Index. This approach follows Greyling and Verhoef (2015), where agricultural output was calculated as the reported volume of the commodities as reported in the censuses, and supplemented the information with export and import values.

In effect the weights are only required to be relative value shares - i.e. weights relative to the other commodities in the larger commodity groups, which are easier to weight. For instance, the livestock production numbers from the census may be compared to each other to form the livestock index. However, only a portion of livestock produced would be sold in any given period, which makes it difficult to compare these numbers to other commodities produced, e.g. the gallons of milk produced. This is also the reason that the products included in the indices for provisions are kept separate from the other products, e.g. wheat and wheat flour are included in separate commodity group indices.

The individual commodities included in each of the 7 commodity group indices are reported in Table 3. The categorisation are informed by the classification of the commodities in the sources of the price information. For instance, in the Blue Books, the prices are divided into Agricultural Produce, Stock and Animal Productions, and Provisions.[12]

Table 3: Commodity classification

Crops	Agri.Produce	Pastoral.Products	Livestock	Pastoral.Provisions	Agri.Provisions	Other.Prov
Wheat	Tobacco	Wool	Cattle	Beef	Bread	Tea
Mealies	Dried Fruit	Hides	Horses Mules & Asses	Mutton	Flour	Coffee
Barley	Wine	Skins	Sheep	Pork	Mealie Meal	Sugar
Oats	Brandy	Cheese	Pigs	Eggs	Boer Meal	Beer
Oathay		Fat & Tallow	Goats	Butter	Oatmeal	Rice
Rye		Soap	Fowls & Ducks	Milk		Salt
Peas & Beans						Candles
Potatoes						

The final step is then to aggregate these 7 commodity group indices together to form an overall index. The difficulty is that the weights are not all comparable, given that some of the commodity groups are weighted with reference to value in the 1904 census, while others are weighted with reference to import value shares. Specifically, import value share were used to create the indices for the three indices of provisions. This makes it difficult to aggregate the indices into an overall wholesale index.

There will always be some difficulties in finding the best weights with historical data like this and different authors have used different assumptions to created relative value shares. For instance, Klovland (2013) computed the relative weights for each commodity within the group as follows: The commodity with the greatest market value was given a load of 10 and the other commodities are scaled proportionately, using rounded integer load values, subject to the constraint that all the time series for which there are data get a load factor of at least one.

In order to combine these 7 commodity groups into an overall wholesale index, a similar assumption is made about the relative weights of the commodity groups as follows: Crops (5), Agricultural Produce (5), Pastoral Products (3), Livestock (8), Pastoral Provisions (5), Agricultural Provisions (3), and Other Provisions (3). Output values of domestic goods are not supposed to be estimated with much precision, the intention is merely to obtain a reasonable set of weights for the construction of the indices. For instance, Klovland (2013) found that in some cases little relevant information was available and figures had to be stipulated on the basis of export shares, reasonable growth rates between benchmark years as well as pure guesswork.[13]

[12] A few commodities had to be excluded from the analysis because of a lack of observations. These include lucerne hay and oranges from the Agricultural Journals, as well pumpkins, aloes, argol, condensed milk and lamp oil from the Blue Books.

[13] Klovland (2014b) makes to further points. Firstly, a case can be made for making some reduction in weights based on gross output measures if a raw material is used extensively in the production of a more finished good, and both are included in the index. In practice it is a matter of judgement how far this principle can be pushed.23 A second point concerns the fraction of agricultural goods that was not sold on the market but consumed on the farms. For historical price indices this may be a highly relevant consideration.

References

Bailey, M. J., Muth, R. F. and Nourse, H. O. (1963a) 'A regression method for real estate price index construction', *Journal of the American Statistical Association*, 58(304), pp. 933–942. doi: 10.1080/01621459.1963.10480679.

Bailey, M. J., Muth, R. F. and Nourse, H. O. (1963b) 'A regression method for real estate price index construction'. doi: 10.1080/01621459.1963.10480679.

Boshoff, W. H. and Fourie, J. (2015) 'When did globalization begin in South Africa?', *Stellenbosch Economic Working Papers: 10/15*, pp. 1–16.

Case, K. E. and Shiller, R. J. (1987) 'Prices of single-family homes since 1970: new indexes for four cities', *New England Economic Review*, (September), pp. 45–56. doi: 10.3386/w2393.

Chen, B. (2007) 'An Empirical Comparison of Methods for Temporal Distribution and Interpolation at the National Accounts', *Bureau of Economic Analysis*.

De Kiewiet, C. W. (1941) *A history of South Africa, social & economic*. Oxford: Clarendon Press.

De Kock, M. H. (1924) *Selected subjects in the economic history of South Africa*. Cape Town, South Africa: Juta.

Fourie, J. (2016) 'The Data Revolution in African Economic History', *Journal of Interdisciplinary History*, XLVII(2), pp. 193–212. doi: 10.1162/JINH.

Giffen, R. (1879) 'On the Fall of Prices of Commodities in Recent Years Source', *Journal of the Statistical Society of London*, 42(1), pp. 36–78.

Gilbert, D. W. (1933) 'The Economic Effects of the Gold Discoveries Upon South Africa: 1886-1910', *The Quarterly Journal of Economics*, 47(4), pp. 553–597.

Greyling, L. and Verhoef, G. (2015) 'Slow growth, supply shocks and structural change: The GDP of the Cape Colony in the late nineteenth century', *Economic History of Developing Regions*, 30(1), pp. 23–43. doi: 10.1080/20780389.2015.1012711.

Hansen, J. (2009) 'Australian house prices: A comparison of hedonic and repeat-sales measures', *Economic Record*, 85, pp. 132–145. doi: 10.1111/j.1475-4932.2009.00544.x.

Jevons, W. S. (1865) 'On the Variation of Prices and the Value of the Currency since 1782', *Journal of the Statistical Society of London*, 28(2), pp. 294–320.

Klovland, J. T. (2013) 'Contributions to a history of prices in Norway: Monthly price indices, 1777-1920', *Norges Bank Working Paper*, (23).

Klovland, J. T. (2014a) 'Challenges for the construction of historical price indices: The case of Norway, 1777-1920', *Norges Bank Bicentenary Project*, (March).

Klovland, J. T. (2014b) 'New methods for construction of historical price indices, with an illustration from Norway, 1777-1920', *European Review of Economic History*, 18(3), pp. 277–305. doi: 10.1093/ereh/heu012.

Kräussl, R. and Lee, J. (2010) 'Art as an Investment: the Top 500 Artists', *Business*, 31(February), pp. 1–26.

Litterman, R. B. (1983) 'A Random Walk, Markov Model For The Distribution of Time Series', *Journal of Business & Economic Statistics*. Federal Reserve Bank of Minneapolis, 1(2), pp. 169–173.

Persons, W. M. and Coyle, E. S. (1921) 'A Commodity Price Index of Business Cycles', *The Review of Economics and Statistics*, 3(11), pp. 353–369.

Sax, C. and Steiner, P. (2013) 'Temporal Disaggregation of Time Series', *The R Journal*, 5(2), pp. 80–87. Available at: https://journal.r-project.org/archive/2013-2/sax-steiner.pdf.

Schumann, C. G. W. (1938) *Structural changes and business cycles in South Africa, 1806-1936*. New York: Staples Press.

Silberling, N. J. (1923) 'British Prices and Business Cycles, 1779-1850', *The Review of Economic Statistics*, V(2), pp. 223–235.

Solar, P. M. and Klovland, J. T. (2011) 'New series for agricultural prices in London, 1770-1914', *Economic History Review*, 64(1), pp. 72–87. doi: 10.1111/j.1468-0289.2009.00527.x.

Wang, F. T. and Zorn, P. M. (1997) 'Estimating house price growth with repeat sales data: What's the aim of the game?', *Journal of Housing Economics*, 6(1997), pp. 93–118. doi: 10.1006/jhec.1997.0209.

Zwart, P. de (2011) 'South African Living Standards in Global Perspective, 1835-1910.', *Economic History of Developing Regions*, 26(1), pp. 49–74. doi: 10.1080/20780389.2011.583003.