Limits and Uses of Price Tests for Market Definition

WILLEM H. BOSHOFF

REVISED JUNE 2013

Stellenbosch Economic Working Papers: 01/11

KEYWORDS: MARKET DEFINITION; PRICE CORRELATION; UNIT ROOT; BOUNDS TEST; LAW OF ONE PRICE
JEL: L40, L11, C32, D4

WILLEM H. BOSHOFF
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: WIMPIE2@SUN.AC.ZA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Limits and Uses of Price Tests for Market Definition

WILLEM H. BOSHOFF

ABSTRACT

Various quantitative tools have been developed for defining markets in competition law investigations. Econometric tests on price co-movement represent one such set of tools: two regions or products are considered part of the same market if their prices co-move. However, price co-movement tests, especially the more advanced econometric tests, have been criticized in the competition policy literature. Critics claim that price tests focus only on price linkages and, furthermore, that some of the advanced price tests are misleading in small samples. This paper applies a range of price tests, including correlation analysis, Granger-causality tests, unit root tests and the recent autoregressive distributed lag (ARDL) bounds test, to data from the 2006-2008 competition investigation into business practices in the South African dairy industry. We argue that the different price tests ask different questions and that it is not useful to dismiss an advanced price test if it suggests a different market than that identified by a simple correlation statistic. We also consider the criticism of poor small-sample performance of price tests: many conventional tests have long been shown to suffer from small-sample power and size problems, but critics fail to account for recent improvements in this regard. The paper concludes that the combination of various price-test results offers a rich picture useful for market definition purposes, especially if they are employed as exploratory tools rather than confirmatory ones.

Keywords: market definition; price correlation; unit root; bounds test; law of one price
JEL codes: L40, L11, C32, D4

Price elasticity estimates are the preferred tools for market definition in competition law investigations, as these estimates provide direct evidence of market power. However, data and time constraints in competition investigations frequently prevent the estimation of price elasticity and force practitioners to rely on less sophisticated tools. One such set of tools are time-series tests of price co-movement, which are based on the law of one price (LOOP) and consider the extent to which prices in different regions or for different products are related. While considered useful by practitioners, price tests are often criticised on grounds that they do not accord with markets for antitrust purposes, are inconsistent with another, and perform poorly for short time series. More important, their results can be affected by common shocks influencing both price series. Without deeper institutional knowledge about the market in question, price tests can therefore lead to spurious market definitions.

This paper explores the limits *and* uses of price tests as tools for market definition in South African competition policy. We build on the empirical example in Boshoff (2007), which deals with the definition of South African milk markets at the producer-processor level. Boshoff (2007) and Mncube, Khumalo, Mokolo, and Nijisane (2008) introduce price tests in the South African context. Boshoff, in particular, considers the application of a specific type of price test and suggests a conceptual framework where price tests are used as confirmatory tools for a market definition hypothesis. Subsequent work by Muzata, Robb and Maphwanya (2012) criticises the Boshoff findings, noting alternative qualitative evidence and considering the limits of focusing on price relationships for the milk market case. These authors also criticise the performance of price tests for two other competition cases. It is therefore clear that the use of price tests is a subject of debate among South African competition practitioners and scholars. This paper contributes to the debate in three ways. Firstly, it discusses the merits of using price tests as part of a larger toolkit for market definition, addressing the specific conceptual and empirical criticisms raised in the literature. Secondly, the paper explores the full range of price tests available, discussing briefly the unique focus of each test and suggesting new tests that may perform better than conventional price tests. Thirdly, the paper applies the full range of price tests to the Boshoff data, noting the problems and limitations of the Boshoff analysis and discussing the implications for market definition.

1. Price tests and their problems

Stigler and Sherwin (1985) were among the first to formally investigate the use of the law of one price (LOOP) in delineating markets for competition policy purposes. The LOOP, articulated as early as the late nineteenth century (see Cournot (1927) and Marshall (1920)), posits that prices within a single market should converge, allowing for some variability related to transport costs and exchange rates, among others. Authors applying the LOOP generally argue for relative price convergence rather than for the more stringent absolute price convergence as the condition for market singularity: prices in two regions of the same market need not be equal, but should be related (Haldrup, 2003). Furthermore, the LOOP literature emphasises that a study of relative price convergence should be sensitive to the particular context, as factors such as transaction costs, degree of product differentiation and the time required for prices to adjust to arbitrage can affect conclusions regarding relative price convergence (see Hunter (2008:68-73) for a summary of the LOOP literature). However, regardless of the type of convergence, many academics and practitioners are sceptical of the use of tests of price convergence for market definition in competition cases. These critics argue that a market for competition policy purposes cannot be identified from price relationships alone and raise a number of conceptual and empirical criticisms of price tests.

1.1 *Conceptual problems*

Critics argue that price co-movement tests only establish whether price series in different regions are ‘linked’, but do not verify whether firms have the capacity to raise prices (Massey, 2000: 317-318). Consequently, price elasticity is argued to be the only appropriate measure for market definition in competition law investigations. Hosken and Taylor (2004), Genesove (2004), and Massey (2000) argue that geographic market definition based on price co-movement, under very general conditions, could be misleading and that institutional knowledge of the relevant market is a prerequisite for correctly interpreting price test results. Along similar lines, McCarthy and Thomas (2003: 15) point to cases where two regions exhibit significant price co-movement, but supply constraints prevent producers in one region from competing with producers in the other region. They argue alternatively that regions for which price co-movement is not substantial may very well constitute a single market, if one of the regions holds excess production capacity.

Price tests need not be inconsistent with the concept of an antitrust market. In competition policy – at least if one relies on the hypothetical monopolist test – a market is the smallest possible space containing the smallest possible set of close substitutes that would, if supplied by a single firm, enable that firm to exercise pricing power. Arguably, then, a tool used for defining markets should be aimed at ranking the various substitute products and identifying from this ranking the closest substitutes. The price series of two products that are close substitutes are likely to be related. In fact, one may even argue that the LOOP must hold for a set of substitutes if these substitutes are to constitute an antitrust market (Hunter, 2008). At worst, conventional price tests (read together with other relevant evidence) may identify a larger market than necessary. Yet newer price tests – that have received little attention from competition policy practitioners so far – may do better as they focus on the *size* rather than only the existence of price relationships.

While price tests may have a relation to market definition, one could argue that price tests are too often misleading if their results are not considered in conjunction with other evidence. This would imply that price tests embody too limited a set of information to be useful in market definition. Yet requiring any quantitative tool to offer conclusive results for market definition (or any competition question) would set too high a standard. The problem here lies with viewing price tests for market definition as confirmatory tools: in this view, price tests generate conclusive quantitative proof on market boundaries to support the intuition provided by anecdotal evidence. Boshoff (2007) also advances this view in the South African context, suggesting that price tests are effectively hypothesis tests. But this is a problematic view, as it ignores the motivation for the use of price tests in the first place – the inability to formally estimate an econometric model and to obtain estimates of price elasticities. The lack of information forces the analyst to gather information and build a consistent story *in a piecemeal fashion* by combining different pieces of evidence. Under such an approach, price tests, rather than being confirmatory, become exploratory tools that help the analyst to weave the story. The signal from these exploratory tools is likely to be imperfect as the tools only consider a subset of information. To require imperfect tools to be accurate would be demanding too much. In fact, one may argue that even an econometric estimate of cross-price elasticity will produce an imperfect signal, as it is derived from an econometric specification imposing a number of assumptions and ignoring a large amount of relevant market information. All statistical tests involve the risk of incorrect inferences and it is up to the analyst to minimise this risk by requiring the results of particular statistical tests to be supported by other qualitative evidence or even, where feasible, alternative statistical tests.

1.2 *Empirical problems*

A further critique of price tests is that they are often unable to discriminate *empirically* between close and less close substitutes. This is best summarised in recent work by Coe and Krause (2008), who use a mainstream product differentiation model to generate price data for three products, of which two are defined close substitutes. The authors apply conventional price tests to the simulated data and show that a number of conventional price tests have difficulty in identifying a single market for the two close substitutes and a separate market for the third. They therefore conclude that price tests are less useful for market definition.

There are limits to the Coe and Krause arguments, however. Firstly, it is not a good strategy to demonstrate the poor performance of quantitative tools by using idiosyncratic scenarios that do not often occur in practice. For more realistic market scenarios, the price tests appear to discriminate well between close and less close substitutes. Of course, given the limited and noisy data often encountered in practice, such an assessment of the price test results may well require the analyst to allow for lower levels of statistical certainty than the conventional 1% or 5% confidence levels. But, at the very least, these test results would indicate crude signals concerning price convergence.

Secondly, this negative assessment of advanced price tests is often due to older versions of price tests with low statistical power and size. As argued in the empirical application, it is likely that at least some of the negative conclusions regarding the usefulness of price tests would be overturned if newer tests for unit roots or equilibrium relationships are used. Another explanation, and one that receives significant attention in this paper, is that analysts often do not appreciate that price tests ask different questions. While newer price tests aim to address the statistical problems experienced by earlier versions, they sometimes ask different or more specific questions than the original tests. A battery of different tests may therefore provide different results that are nevertheless mutually consistent. It is therefore problematic to argue that tests for cointegration relationships should provide conclusions similar to those of correlation analysis.

Price tests may therefore be useful, also in the South African context, provided that competition analysts understand their limitations, as discussed in the following section.

1.3 *Implications for South African competition policy*

The criticisms discussed above suggest that price tests should not be oversold as tools for market definition. In his application of price tests to a South African market definition problem, Boshoff (2007) presents price tests as confirmatory tools that can ultimately determine the judgment about market boundaries. Based on the preceding discussion, this is an overoptimistic assessment. Practical constraints force us often to rely on the piecemeal gathering of evidence, from which it is difficult to discern a particular hierarchy of evidence. It is therefore appropriate to consider a variety of evidence, including price tests, which can offer a systematic analysis of an important dimension of competitive behaviour.

This approach to the use of quantitative tools is not unfamiliar to competition practitioners. Consider, for example, the use of time-series techniques in detecting cartels. There is an entire literature on the use of cartel screening tools – tools that are, by nature, limited in information but can be used in the first stage of competition investigations (see Abrantes-Metz (2012) for an accessible introduction). In a similar way, price tests may fit in well with the initial stages of a competition investigation, when the analyst is building an understanding of how the particular market (and related markets) functions.

If price tests are to be used, it is important to be aware of their small-sample power: new price tests with better power and size properties (discussed later in this paper) can be useful to South African practitioners who often work with small data sets with an insufficient time span or data frequency (also see Katsoulacos, Konstantakopoulou, Metsiou and Tsionas (2012) for an optimistic assessment of the small-sample power of price tests). Furthermore, a variety of price tests are available, each focusing on a particular dimension of a price relationship. Boshoff (2007) focuses on unit root tests (discussed below), but undersells other price tests as well as the need for a range of price tests. As argued above, different price tests are not intended to confirm one another, but to provide alternative perspectives on the nature of price relationships. It is therefore important to consider, in greater detail, the range of price tests. The following section presents these tests, discussing their link to market definition and their individual limitations and uses.

2. Tests of price co-movement

Price tests can be divided into two groups: one group focusing on short-run relationships and another on long-run relationships. Correlation statistics and Granger-causality tests fall into the first group, while unit root tests, cointegration tests and autoregressive distributed lag models fall into the second group. A 'long run' relationship between variables refers here to an equilibrium relationship between the variables. The time horizon of the 'long run' in any particular case is an empirical matter and one way to measure the 'long run' time horizon is by the speed with which the variables adjust following a disturbance of the equilibrium among them.

2.1 *Correlation statistics*

The simple correlation statistic is a popular tool for assessing short-run price relationships between products or regions. Even so, correlation statistics as tools for market definition face two major challenges. Firstly, they lack an objective benchmark against which to assess their economic significance, as it is difficult to decide whether a particular correlation statistic is economically meaningful (Forni, 2004: 450). For example, it is not clear whether a correlation statistic of 0.5 between two product price series suggests a meaningful substitution relationship for market definition purposes. One solution may be to compare correlation statistics between two variables with correlation statistics for products known to be substitutes (Davis and Garcés, 2010) or to similar correlation statistics in other markets where market integration is known (Bishop and Walker, 2002). Nevertheless, these benchmarks remain problematic. Secondly, correlation analysis for market definition purposes is frequently employed without regard to whether the calculated sample value can be generalised to a larger population, i.e. no regard is paid to statistical significance. This may be especially problematic in small samples – a frequent occurrence in competition analysis. Furthermore, notwithstanding these challenges, it is difficult to arrive at a final summary of relationships, especially if the response of one price variable to another is not immediate but protracted. For this reason, regression-based price tests have been developed.

2.2 *Granger-causality tests*

Granger-causality tests for market definition gained momentum in the 1980s (as summarised in Bishop and Walker (2002)) and were among the first competition analysis tools to involve regression analysis. A price series $p_{1,t}$ is said to 'Granger-cause' another price series $p_{2,t}$ if the past and present values of $p_{1,t}$ provide information to forecast future values for $p_{2,t}$ (Granger, 1969). The tests are based on the notion that if the LOOP holds across two regions, price disturbances in one region should translate into price disturbances in another region; prices in one region or of one product should Granger-cause prices in another region or another product if the two products or regions share a common market.

While these tests give an impression of sophistication, they face three significant challenges when used to define markets. Firstly, Granger-causality tests focus on the existence rather than the size of a relationship. In this sense, the test overcomes the problem of an objective benchmark faced by correlation analysis. But for a market definition test, the issue of concern is not only whether a statistically significant relationship between two price series exists, but also whether two price series are *meaningfully* related to the extent that they will pass the hypothetical monopolist test (Bishop and Walker, 1998: 451-452). Secondly, Granger-causality tests impose a one-way logic on market definition. By nature, these tests are concerned with questions of direction: does $p_{1,t}$ predict $p_{2,t}$, or the other way around, or does it perhaps work both ways? While one need not focus on direction per se, it is not clear whether a two-way finding should be interpreted as stronger evidence of market integration compared with a one-way finding. Thirdly, Granger-causality tests (like correlation analysis) focus exclusively on short-run relationships (Pagan, 1989). In fact, the tests require price data that are stationary – and the process of first-differencing data to obtain such stationary data involves a significant loss of long-run information. Of course, one may prefer a focus on short-run relationships for market definition purposes, but, even then, the regression models have to properly account for long-run parameters if the short-run parameters are not to be biased. This particular concern can be addressed by considering tests for long-run equilibrium relationships, including stationarity tests and cointegration tests.

2.3 *Stationarity tests*

The past twenty years has seen the use of a number of tests for long-run equilibrium, focused on the concept of cointegration (Engle and Granger, 1987). Two non-stationary series are said to be co-integrated if a linear combination (known as the cointegrating relationship) of the series is stationary.

Cointegration tests establish the existence of a cointegrating relationship, and affirmative proof from these tests can be taken as evidence for a single market. However, Forni (2004) argues that it is prudent to test whether the cointegrating relationship actually takes the form $(1; -1)$, as such a one-to-one relationship indicates a perfectly integrated market. Finding a one-to-one relationship between prices in two regions clearly constitutes strong evidence that the two regions form a single market, but is not a necessary condition for market singularity. A more general cointegrating relationship of the form $(1; -\beta)$, for any real-valued β, also indicates market singularity, as prices are still related – even though prices only partially converge and are more difficult to interpret.

Nevertheless, a test for a co-integrating relationship of $(1; -1)$ has practical use in competition policy settings as such a test can be performed without a formal cointegration analysis. This follows because of the equivalence between a test for a cointegrating relationship of $(1; -1)$ and a test of whether the log price ratio for the two regions is stationary (Forni, 2004). Therefore, in practical terms, a test for the existence of a $(1; -1)$ cointegrating relationship can be done by calculating the log price ratio for the two regions and then testing whether this ratio is stationary, using a standard unit root test.

Boshoff (2007) and Mncube, Khumalo, Mokolo, and Nijisane (2008) discuss the application of stationarity tests to South African market definition problems. Boshoff also summarises the main limitations of unit root tests as tools for market definition. More recently, Coe and Krause (2008) criticised, among others, unit root tests as tools for market definition. Coe and Krause argue that unit root tests provide the incorrect market definition signal when applied to simulated price data for three products generated from an economic model in which two of the products are close substitutes. However, as is argued below, the Coe and Krause results may be the result of the low small-sample power of their unit root tests – an issue that can easily be ameliorated by using improvements suggested by the econometrics literature over the last decade. In fact, Boshoff (2007) discuss the main univariate tests, including the four tests suggested by Ng and Perron (2001), as well as more recent panel unit root tests with greater statistical power and better size properties in small samples.

Nevertheless, it is important to acknowledge that unit root tests may be limited for practical market definition purposes as the available price series in competition investigations are frequently of a fairly short time span, rarely more than five years. A substantial literature, initiated by Perron (1991), shows that test statistics for unit roots are consistent only when the time span increases with the number of observations (Maddala and Kim, 1998). Therefore, even if one obtains price series of a high frequency (say monthly), unit root tests may still be problematic. Unit root tests generally require enough data points over which mean reversion could occur in order to distinguish between series with and without stochastic trends. Nevertheless, there is no absolute minimum, as mean reversion and behaviour may depend on the particular market, and it is still possible to achieve a sufficient amount of mean reversion over a relatively short span.

2.4 *Cointegration tests and ARDL bounds tests for level relationships*

Stationarity tests offer one way of testing for a cointegrating relationship without performing a formal cointegration analysis. These tests, however, are quite specific: they are strict tests for both the existence of a long-run relationship and, if present, for a one-to-one long-run relationship. Analysts interested in allowing for a more general set of long-run relationships are forced to perform cointegration tests.

The standard approach to testing for cointegration is the Johansen systems approach, which studies the properties of the long-run matrix of a vector autoregression of the non-stationary variables. The Johansen approach has greater power than the Engle-Granger method, which is the original (single-equation) approach to cointegration testing. Yet even the Johansen tests are exposed to potential unit root pretesting bias – given that the non-stationarity of the price series have to be verified prior to performing the cointegration analysis. The implications of an incorrect inference on unit root properties are substantial, particularly in smaller samples – this probably limits the use of cointegration analysis in a market definition context where small samples are common. While panel unit root tests with higher power can reduce this bias, cointegration tests themselves also suffer from small-sample power problems.

An alternative approach to testing for cointegration – and one that does not depend on prior unit root testing – is the ARDL bounds testing approach developed by Pesaran, Shin and Smith (2001) (PSS). The bounds test aims to test for the presence of a long-run relationship *regardless of the order of integration of input variables*. In other words, the ARDL bounds test is a cointegration test when the input variables are non-stationary, but can also be applied to test for long-run relationships between stationary variables. This implies that the bounds test also allows for a broader class of long-run relationships to qualify for market definition purposes.

Despite econometric advances, long-run tests may not be the preferred tool for market definition. In the long run, most consumers have the capacity to respond to price changes, so that relying on long-run equilibrium is likely to overstate the size of the market. Nevertheless, short-run analysis may still benefit from long-run analysis. Cointegration can assist in building an error-correction model (ECM), by adding long-run relations (represented in error-correction term z_{t-1}) to the following system of short-run dynamics:

$$\Delta p_{1,t} = \gamma_1 z_{t-1} + lagged(\Delta p_{1,t}, \Delta p_{2,t}) + \varepsilon_{1,t}$$

$$\Delta p_{2,t} = \gamma_2 z_{t-1} + lagged\left(\Delta p_{1,t}, \Delta p_{2,t}\right) + \varepsilon_{2,t}$$

Without the z_{t-1} term, the above system would be misspecified. The inclusion of the term therefore allows a more accurate estimation of short-run dynamics (represented as $lagged\left(\Delta p_{1,t}, \Delta p_{2,t}\right)$). The coefficients in $lagged\left(\Delta p_{1,t}, \Delta p_{2,t}\right)$ indicate the degree to which short-run movements in the two price series are related and can be used to define the relevant market. The coefficients γ_1 and γ_2 then indicate the speed of adjustment and, hence, the 'horizon' of the relationship between the price series.

The bounds test is based on a conditional ECM, and the test statistic is the F-test for the statistical significance of lagged variables in *levels* in this ECM. PSS derive asymptotic distributions for the test statistic that assume either all variables non-stationary or all variables stationary. These two sets form the upper and lower bounds of the test. If the test statistic exceeds the upper bound or falls below the lower bound, the inference is straightforward, but the outcome is unknown when the test statistic falls between the two bounds. In the latter case, it is still necessary to test for stationarity.

The ARDL bounds test is easy to implement, which further recommends it for use in market definition. As discussed in Appendix A, the procedure involves a single-equation estimation using OLS and a straightforward testing of coefficient restrictions using the F-test. Consequently, practitioners with only an introductory technical knowledge of econometrics can perform bounds tests.

2.5 *Summary*

The preceding sections discuss a number of tests for price relationships in the short run (correlation statistics and Granger-causality tests) and in the long run (unit root tests, cointegration tests and the ARDL bounds test). Each test is shown to ask a unique question concerning a price relationship, i.e. each test proceeds from a specific null hypothesis. Differences in null hypotheses are important if one is to compare the results of various price tests. Critics often argue that the various price tests are inconsistent with one another and therefore not suitable tools for market definition. But such an argument assumes that the various tests focus on the same question, which they clearly do not. Nevertheless, results for short- and long-run relationships may well differ and it is important to be aware of the exact nature of price relationships when defining markets.

As argued earlier, Boshoff (2007) promotes the use of unit root tests for the definition of milk markets in the South African context. It would be interesting to apply the full battery of price tests discussed above to the milk price data, in order to consider the robustness of the market definition suggested by the price test (unit root tests) employed in Boshoff (2007). Such an empirical application would help to illustrate how price tests are to be applied in practice, taking consideration of the heterogeneity in focus among the various tests.

3. Case description and existing evidence on the relevant market

Based on the preceding assessment, this paper applies a range of price tests to data from the South African milk case described in Boshoff (2007). Both Boshoff (2007) and Maphwanya, Muzata and Robb (2012) provide a detailed description of the qualitative evidence, and this section provides only a brief description of the case and qualitative evidence as well as a summary of the existing research by Boshoff and others on the relevant market in this case.

3.1 *Case description*

The competition investigation focused on exclusive supply agreements concluded between dairy processors and dairy farmers belonging to SAMILCO, an industry body representing South African dairy farmer interests. The agreements required members of SAMILCO to sell all milk production to the processor, or risk losing membership of SAMILCO. Agreements were concluded for a three-year period, after which they would continue but with the option of terminating after a six month notice period. The competition authority held that these agreements constituted anti-competitive vertical restraints. The competition investigation also covered a range of other practices, including coordination among dairy processors and price fixing, but the focus here is on the market definition question as it relates to the vertical restraints issue (see Competition Tribunal (2008: 3) for a summary of the allegations).

As discussed earlier, the milk case is a particularly relevant case because of the debate surrounding the use of price tests of milk data. A major limiting factor is that the case was never concluded, so that there is no formal market definition identified by a court against which to test the empirical conclusions. Nevertheless, as we argue earlier, it is important to explore the various dimensions of price tests and the consistency of the various tests in order to provide guidance to practitioners.

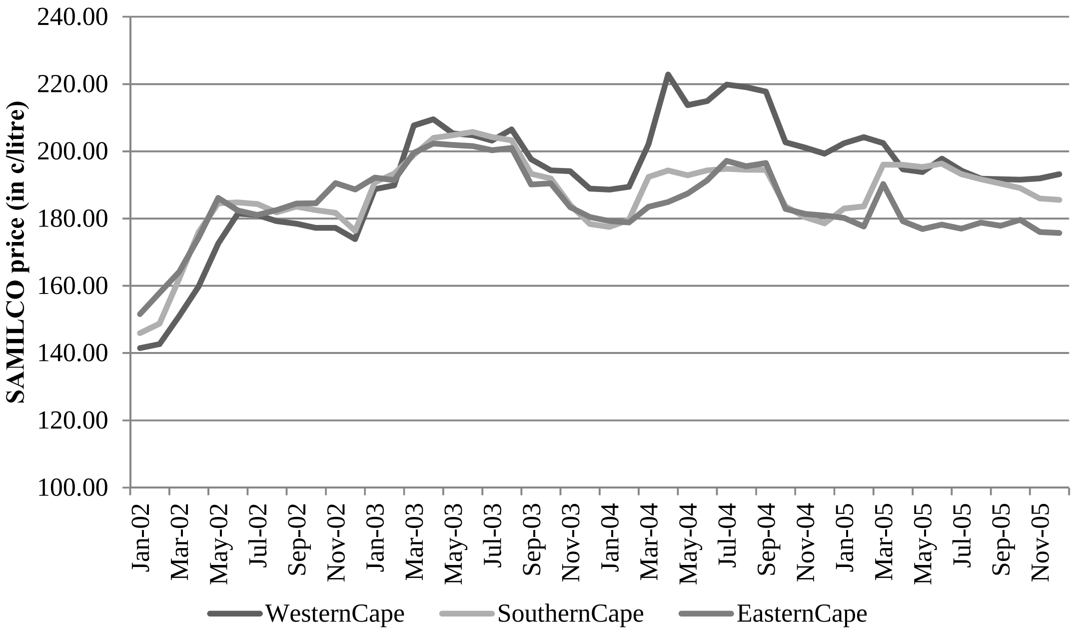
3.2 *Previous findings on the geographic market*

Given that the alleged agreements were concluded upstream between dairy processors and farmers, but may have had an effect on final downstream consumers, two product markets are relevant: an upstream market for fresh milk between dairy processors and dairy farmers as well as a downstream market between dairy processors and final consumers (via retailers). In this case, the important market definition question is the geographic scope of the upstream market. One of the dairy processors investigated owned processing plants across three southern milk regions (labelled as Western Cape, Southern Cape[1] and Eastern Cape). For the purpose of assessing whether this particular processor was dominant, it was necessary to establish whether the adjacent southern regions constitute a single market[2].

Boshoff (2007) presents descriptive evidence on geographic substitutability in South African milk markets before proceeding to unit root tests. This paper does not repeat this discussion, which focused on transport cost of milk among the different plants. Boshoff interprets the evidence as suggestive of low transport costs and a single geographic market encompassing the southern plants and, perhaps, even plants in the north. In contrast, more recent work by Muzata, Robb and Maphwanya (2012) presents evidence that support an alternative, narrower, definition of the geographic market.

Boshoff explores the single market hypothesis by considering results from various unit root tests of price ratios, performed on milk price data from an agricultural co-operative, the Southern Africa Milk Co-operative (SAMILCO) (reported in Figure 1).

Figure 1: SAMILCO prices in Western Cape, Southern Cape and Eastern Cape, January 2002 – December 2005



[1] These are coastal regions located in southwest and southern South Africa and not provinces.

[2] In the current case, the focus falls on the use of buyer power. The HM test is articulated in terms of *seller* power (in other words, the market is defined by considering the response of the processor to a 'small but significant non-transitory increase in price' by the dairy farmer). However, one may also view the price tests as an implementation of an SSNDP (small but significant non-transitory decrease in price) by a buyer.

SAMILCO represents a large portion of dairy farmers in the southern regions and calculates an average monthly milk price for each of the Western Cape, Southern Cape and Eastern Cape regions. Figure 1 suggests changes in the distributional properties of the individual SAMILCO price series, with the series moving arbitrarily far from a deterministic trend. All three series are non-stationary (analyses available on request) and are seasonally adjusted.

Here it becomes important to take note of institutional changes in the market. Prior to 2005, the market relations between dairy processors and farmers were governed by supply agreements, which influence prices, but prices in different regions were not contractually linked. Nevertheless, market prices in the different regions generally co-moved until January 2004, when the prices of Western Cape milk increased significantly following very strong actual and expected increases in milk demand. Prices also rose in the Southern and Eastern Cape but to a lesser extent. Southern Cape farmers therefore demanded higher prices from dairy processors, failing which farmers signalled that they would use the arbitrage opportunity to sell their product in the Western Cape. This resulted in a determined effort to match prices in the two regions from the start of 2005, which culminated in a renewed agreement in December 2005 between the processor under investigation and its Southern Cape suppliers. This agreement explicitly linked the average milk price in the Western Cape with the average milk price in the Southern Cape, allowing for transport costs between the two regions. One could argue that the 2005 changes altered the market dynamics and artificially created a single market across the two regions. Alternatively one could argue that these changes are a reflection of interaction in a market that has always included both regions. Either way, it is necessary to consider two sample periods, one including and one excluding the 2005 data.

Given the institutional background, Boshoff (2007) performs unit root tests on the price ratios and finds evidence on long-run price relationships between the Western and Eastern Cape and Southern and Eastern Cape, but none for the Western and Southern Cape. The panel unit root tests also suggest strong evidence of stationarity of the ratio of prices in each of the three regions and an average price. Therefore, Boshoff argues in favour of a single market.

This conclusion may be too optimistic, given that the unit root tests consider only a very specific type of price relationships over a specific time horizon. Furthermore, the small sample size can create power and size problems – even for the panel versions of the unit root tests. In the following section, we therefore consider the results for a range of price tests applied to the milk price data – showing that the conclusions are not uniform and depend *inter alia* on which time horizon the analyst considers important.

4. Price test results

4.1 *Short-run relationships: correlation and Granger-causality results*

Table 1 reports the contemporaneous correlation among the three price series (seasonally adjusted) for the sample period 2002-2004 (and for 2002-2005 in brackets). Correlation is about 0.7 between Eastern and Southern Cape prices and around 0.5 for Western Cape prices, on the one hand, and Southern and Eastern Cape prices, on the other.

Table 1: Correlation of milk prices in first differences for January 2002 to December 2004 (January 2002 to December 2005)

	Western Cape	Southern Cape	Eastern Cape
Western Cape	1.00		
Southern Cape	0.51 (0.50)	1.00	
Eastern Cape	0.48 (0.45)	0.75 (0.72)	1.00

Price changes in one region are unlikely to be transmitted instantaneously to another region and it is therefore useful to also consider lagged correlations. Lagged correlations (actual values available on request) were rarely in excess of 0.4 for any of the price pairs. Also, whereas the correlation between S&E is strongest contemporaneously but then declines sharply, the lagged correlation for the W&S declines more gradually from zero to two lags. Furthermore, correlations for lags in excess two quarters were never significantly different from zero at 95% statistical confidence levels. In other words, the correlation evidence for market integration is less strong when lagged relationships are accounted for.

As argued earlier, correlation calculations for market definition face a number of challenges. Important, the conclusion that a 0.5 correlation is 'important' is arbitrary in the absence of a benchmark. Also, if one relies on contemporaneous correlation statistics to define markets, one would conclude that the evidence supports a single market across the Southern and Eastern Cape but with less strong support for also including the Western Cape. However, if conclusions based on lagged correlations are preferred to contemporaneous correlations, one may conclude in favour of separate markets given that the lagged correlations are much smaller. It is therefore difficult to reconcile contemporaneous and lagged correlation. Granger-causality attempts to overcome these problems.

Table 2 reports the Granger-causality tests for sample period 2002-2004 for the three price pairs W&S, W&E and S&E. The tests are reported for lag lengths of 1 to 6 months. This allows us to check the sensitivity of results for the choice of lag. We follow this practice also for the stationarity tests and ARDL bounds tests. The table reports the size of the F-test statistic as well as the associated p-values.

Table 2: Granger-causality tests between price pairs W&S, W&E and S&E (January 2002-December 2004)

Lag	Southern → Western	Western → Southern	Eastern → Southern	Southern → Eastern	Eastern → Western	Western → Eastern
1	0.63 (0.43)	0.04 (0.85)	0.41 (0.53)	0.66 (0.42)	1.96 (0.17)	0.18 (0.68)
2	2.55* (0.10)	0.06 (0.94)	3.12* (0.06)	0.37 (0.69)	4.92** (0.02)	0.13 (0.88)
3	1.64 (0.21)	0.11 (0.96)	2.40* (0.10)	0.26 (0.85)	3.19** (0.04)	1.66 (0.21)
4	1.66 (0.20)	0.57 (0.69)	1.92 (0.15)	0.38 (0.82)	2.94** (0.04)	0.96 (0.46)
5	0.96 (0.47)	2.84** (0.05)	1.72 (0.18)	0.65 (0.67)	1.91 (0.14)	0.78 (0.58)
6	1.99 (0.13)	1.69 (0.19)	1.50 (0.24)	0.47 (0.82)	1.98 (0.13)	0.74 (0.63)

Note: *** Reject at 1%, ** Reject at 5%, * Reject at 10%

We find some evidence of a Granger-causal relationship between E&W prices and E&S for a lag order of two to three months. This also holds for the 2002-2005 period (results available on request). We find little evidence of Granger-causality between W&S prices – there are some significant results but these have no systematic pattern to support a positive interpretation. Also note that in all of the cases, the Granger-causality results are significant in one direction but not the opposite. As discussed, there is no clear reason for requiring bidirectional or unidirectional causality, which may create further questions.

As argued earlier, a comparison of Granger-causality and correlation results should be sensitive to their differing aims. The Granger-causality test is applied with the aim of confirming the *existence* of a dynamic short-run relationship (ignoring the extent of the relationship). Nevertheless, the test results above suggest a short-run relationship between E&S prices, which is consistent with the strong correlation results obtained earlier. The evidence of a strong contemporaneous correlation between W&E prices also finds some support from the Granger-causality test (albeit much weaker). There are no convincing results for a relationship between W&S prices, which might undermine the suggestion that the

Western Cape formed a single market with the other two regions.

Correlation analysis and Granger-causality tests are based on short-run information. Even if one is only interested in studying short-run relationships, econometric research shows that the exclusion of long-run information from short-run models creates misspecification error. The following subsections consider the results of price tests dealing with long-run relationships.

4.2 *Long-run relationships: price-ratio stationarity tests*

As discussed, a popular test for long-run relationships between prices in any two regions is a stationarity test on the ratio of these prices. This is the approach promoted by Boshoff (2007) when dealing with the milk market problem. Boshoff considers log price ratios for the following pairs of regions:

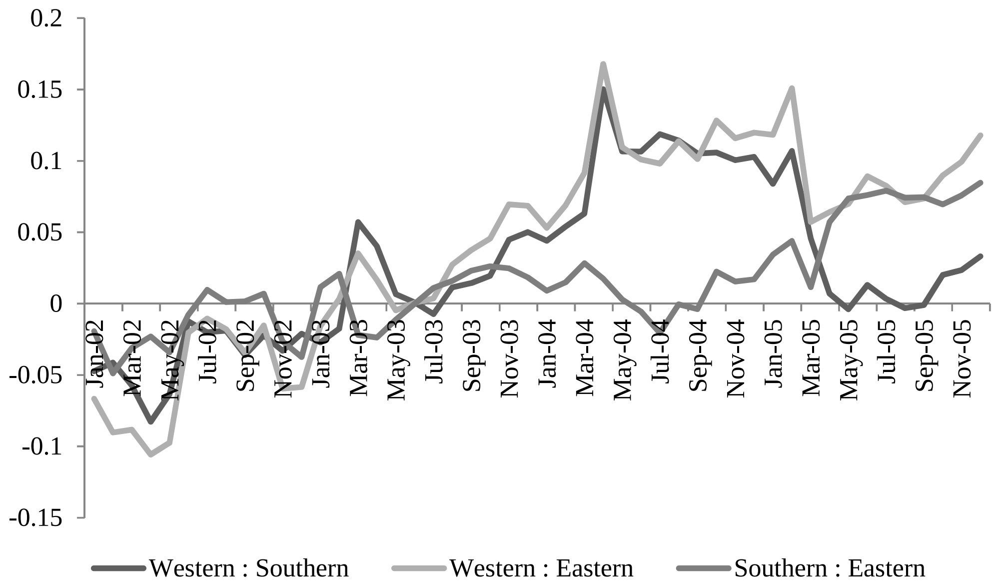
(i) Western Cape and Southern Cape (hereafter called the W:S ratio).

(ii) Western Cape and Eastern Cape (W:E ratio).

(iii) Southern Cape and Eastern Cape (S:E ratio).

Figure 2 reports the three ratios over the sample period including 2005. The W:S graph suggests non-stationarity, due to the strong persistence towards the end of the sample period. Figure 2 confirms the significant changes in 2005 during which Southern Cape prices were brought back in line with Western Cape prices, with a sharp decline in the W:S ratio during this period. The graphical intuition regarding stationarity is not clear for these ratios. Without a deterministic long-run path, both series may have a fairly constant average and variance, indicative of stationary behaviour. However, the wave-like persistence in the S:E ratio may cause one to conclude otherwise.

Figure 2: Log price ratios, January 2002 – December 2005



One can use formal unit root tests to evaluate the stationarity of milk price ratios. As argued earlier, the 2005 institutional changes are likely to have affected price relationships. One should therefore focus on the 2002-2004 sample period. For this period, Boshoff finds evidence of stationarity based on the univariate (Ng and Perron, 2001) tests for two of three price ratios, arguing that this provides sufficient evidence of market integration. Using similar unit root tests, we also find evidence of two stationary ratios (W:E and W:S), but one non-stationary ratio (S:E). The results for the sample period including 2005 suggest non-stationarity, also consistent with the findings of Boshoff.

Boshoff (2007) also suggests, for the purposes of improving power, to study the ratio of prices in each region and an average price calculated over all regions using a panel model. The Levin, Lin and Chu (2002) (LLC) and Im, Pesaran and Shin (2003) (IPS) panel unit root tests on these ratios suggest evidence in favour of stationarity and, therefore, a single market for 2002-2004. Our panel unit root results for 2002 to 2004 are reported in Table 3. We also find strong evidence of stationarity in this period, robust to lag choices up to around four months.

Table 3: Panel unit root test results, January 2002 to December 2004

Lag	LLC	Breitung	IPS	ADF Fisher
1	-5.85*** (0.00)	-4.79*** (0.00)	-3.75*** (0.00)	24.11*** (0.00)
2	-3.94*** (0.00)	-2.45** (0.01)	-2.37** (0.01)	15.44** (0.02)
3	-2.79*** (0.00)	-1.61** (0.05)	11.28* (0.08)	16.83** (0.01)
4	-2.41** (0.01)	-1.31* (0.10)	-1.63** (0.05)	11.59* (0.07)
5	-0.81 (0.21)	-1.55* (0.06)	-0.52 (0.30)	5.34 (0.50)
6	0.12 (0.55)	-1.84** (0.03)	-0.19 (0.42)	3.49 (0.75)

Tests on the sample period including 2005 suggest strong evidence of non-stationarity, also consistent with the earlier Boshoff findings (Appendix B). The Hadri (2000) test, based on the null hypothesis of stationarity (rather than a unit root), confirms that the 2002-2004 price ratios are stationary and non-stationary when including 2005. Therefore, all panel unit root tests appear to agree that price ratios are stationary in the period excluding 2005 and, therefore, suggest a single market.

Table 4: Hadri panel unit root test results

Hadri test statistic	2002-2004	2002-2005
Z-statistic	0.46 (0.32)	4.33 (0.00)
Z-statistic (heteroscedasticity-consistent)	0.60 (0.27)	4.28 (0.00)

While our unit root results reflect those of Boshoff (2007), how does one reconcile these findings with the short-run conclusions? Firstly, if one considers the panel unit root tests indicative of overall stationarity, one would conclude in favour of a single integrated market. This would be a stronger conclusion than was the case for the short-run relationships, which appear to be more nuanced. Secondly, however, as the panel unit root tests do not prove conclusively that all series are stationary (only that some may be stationary), one has to rely on the univariate results. These are more nuanced, as discussed in Boshoff (2007). The stationary finding for the W:E ratio suggests a single market, which is the same conclusion drawn from the short-run findings. In other words, regardless of whether it is long- or short-run price co-movement that matters for market definition, evidence indicate that the Western and Eastern Cape milk regions constituted, at least until the point of institutional intervention, a single market. The non-stationary finding for the S:E ratio suggests separate markets, which is in contrast with the strong signals of market singularity provided by the short-run analyses. In a similar way, the finding of a stationary W:S ratio also contrasts with the earlier analyses, which did not indicate strong short-run relationships between the Western and Southern Cape.

The contrast between short-run and long-run results is one of the primary criticisms of econometric price tests raised by Coe and Krause (2008) and also raised by Maphwanya et al (2012) in their discussion of the milk case. But conflicting results do not necessarily indicate the empirical failure of a particular econometric test. As noted, differences may reflect the different questions that different techniques seek to answer. Nevertheless, a competition analyst should understand the source of these differences. Therefore, it would be important to consider other long-run tests: contrary to the Boshoff (2007) argument, one cannot take the results of the unit root tests as conclusive evidence of long-run relationships. Furthermore, even if one does find a long-run relationship, it may not be particularly important in monthly price changes – i.e. the speed with which a particular price adjusts to a long-run disequilibrium may be so slow as to be negligible and therefore of less importance in a market definition exercise. Therefore, as argued earlier, it is useful to also pursue a technique that combines short- and long-run information and allows for a more general long-run relationship than the one-to-one relationship assumed here.

4.3 *Long-run relationships: cointegration and ARDL bounds tests*

Cointegration tests, similar to unit root tests, have low power in small samples. One solution is to consider panel cointegration tests. Appendix C reports a summary of the results of Pedroni (2001) panel cointegration tests performed on the three price series on the one hand and an average price series across the three regions on the other hand. The results are reported for the shorter and longer sample periods and for various lags to test for robustness. The Pedroni test encompasses seven statistics, four associated with within-variation and three associated based on between-variation. The findings are summarised as supportive of cointegration or not. For the sample period ending in 2004, there is strong confirmation of cointegration for lags 1 to 3, with 4 also suggesting some evidence. After 5 lags, as for the unit root tests, the power of the tests is reduced by the small number of data points. In contrast, and consistent with the unit root findings, there is no evidence of cointegration over the entire sample period. Fisher-type cointegration tests have greater power than the Pedroni cointegration tests. For these tests, we find similar evidence of a unique cointegration relationship, although the finding is less robust to lag choice. Consistent with the Pedroni results, the Fisher-type tests do not provide any evidence of a cointegration relationship among the three milk price series for the sample period including 2005.

Read in conjunction with the unit root results and the institutional knowledge of a structural change at the start of 2005, one could therefore present a strong case, based on long-run relationships, for a single market: both the unit root tests and the cointegration tests suggest evidence of long-run relationships for 2002-2004. The results for 2005 point to no relationship, a result that is likely due to the structural change.

Nevertheless even the Fisher cointegration tests suffer from small-sample power problems, given the very short dataset. Instead, it may be useful to consider an ARDL bounds test for long-run relationships, which avoids potential unit root pre-testing bias and is still applicable when the series are all stationary. While the ARDL method is also a single-equation approach to cointegration (similar to the Engle-Granger approach underlying the Pedroni test), there is evidence that this single-equation approach outperforms the traditional methods.

Table 5 shows the results of the ARDL bounds test for various lags over the two sample periods. We report, firstly, the test statistic, testing the null hypothesis that there is no long-run relationship. Throughout, we find strong support for rejecting this hypothesis – suggesting a significant long-run equilibrium relationship. Secondly, we also report the size of this correlation. The results also suggest a high correlation of around 0.9 throughout. However, thirdly, the speed-of-adjustment coefficient, explaining how quickly monthly milk price changes respond to disturbances in long-run relationships, paints a different picture. In our case the coefficient is around -0.1, which indicates that milk prices take up to 10 months to respond to equilibrium disturbances. This long adjustment period suggests that the long run does not play a major role in the monthly price adjustment process. Consequently, while a long-run relationship may exist, it may not be as important as short-run shocks. This dampens to a significant extent the conclusions of Boshoff (2007), who does not consider the importance of adjustment speeds to long-run equilibrium.

Diagnostic tests identify a number of data outliers, which can be removed by including dummy variables. The inclusion of the outlier dummy variables significantly enhances the model fit. Enhanced model fit for the ARDL models may also explains why, as explained, the bounds test results differ from the unit root test results: unit root tests are less concerned with deriving a congruent representation of the data than the formal ARDL modelling stage that precedes the bounds test.

Therefore, while our results still confirm the cointegration and unit root tests that a long-run relationship is present, it suggests that it may not be economically significant. It may well be that this adjustment behaviour has since changed, but, for the given data in this case, it is too strong a statement to rely on the existence of long-run equilibrium relationships to support one's argument for a single market.

Table 5: ARDL bounds test and long-run results

Lag	**Dimension**	**2002-2004**	**2002-2005**
1	F-statistic	19343.75***	90.39***
2	F-statistic	57.12***	370.22***
3	F-statistic	50.50***	106.91***
4	F-statistic	35.51***	161.19***
5	F-statistic	13.18***	19.28***
6	F-statistic	51.92***	32.87***
1	Long-run correlation	0.83	0.90
2	Long-run correlation	0.96	0.97
3	Long-run correlation	0.85	0.95
4	Long-run correlation	0.87	0.96
5	Long-run correlation	0.99	0.98
6	Long-run correlation	1.20	1.02
1	Speed of adjustment	-0.09	-0.10
2	Speed of adjustment	-0.09	-0.10
3	Speed of adjustment	-0.09	-0.10
4	Speed of adjustment	-0.10	-0.13
5	Speed of adjustment	-0.05	-0.12
6	Speed of adjustment	-0.05	-0.10

4.4 *Summary of findings*

The price test results paint a rich picture of the market. Running a battery of tests adds significant information that aids in a further understanding of milk market boundaries. Short-run tests, based on correlation analysis and Granger-causality tests, support the definition of a single geographic market. Long-run tests, based on unit root and cointegration tests, confirm the existence of equilibrium relationships among the regions, but the ARDL results indicate that these relationships are less important to the dynamic behaviour of milk prices from month to month.

The results confirm the Boshoff (2007) unit root findings, even though these tests consider only a particular dimension of milk price behaviour and are therefore limiting. However, and contrary to Boshoff's claims, the price tests do not offer sufficient evidence of a single market. The analyst has to consider which of the short or long-run relationships should receive greater attention. In this case, if the long run is an important consideration, the analyst would have to explain why a slow adjustment period (10 months) is sufficient to argue that the separate regions exert competitive constraints on each other. It is here where information about transport costs as well as other qualitative information will become important in judging the final boundaries. Ultimately, the price tests have to be triangulated with other evidence. In the current case, there is significant disagreement about this triangulation, as discussed earlier.

Where does this leave price tests as tools for market definition in this case? We would argue that, while price tests are not sufficient in confirming milk market boundaries, they represent a necessary part of the investigation. If competition analysis is about understanding pricing power, an analysis of price relationships seems to us to be a useful starting point. Once evidence of price relationships is established, analysts would have to investigate the source of these relationships. It may be that common demand or supply shocks during the sample period affect all three regions, leading to a spurious conclusion of a single market. Alternatively, demand-switching behaviour may be driving relationships, confirming a single market. Either way, one cannot avoid explaining the milk price relationships before drawing conclusions about the extent of the relevant market.

This brings us to a more general argument about the use of econometric tools in competition investigations. Given their statistical sophistication, one could argue that they should provide final answers to the difficult economic questions facing the parties to a competition investigation. Yet even the most sophisticated econometric tools cannot provide conclusive evidence, given that they all rely on a particular theoretical structure and a particular dataset. It is therefore essential to employ tools that provide some systematic evidence, such as price tests, to be combined with other pieces of evidence when defining markets. Besides, the price tests provide evidence beyond the market definition stage – even if evidence is not strong enough to label two regions as strong competitive constraints, there may well be a number of regions that represent a second-tier of weaker competitive constraints. Understanding these 'fringe' competitors is often important to competition analysts.

From another perspective, one can also see price tests as screens that are useful to employ at the start of the investigation, similar to how time-series econometrics is used in cartel detection. Few analysts would argue that cartel screening tools are always accurate, yet they help the analyst sift through the evidence by suggesting potentially interesting patterns to investigate further. In the current case, there is perhaps fewer data, but competition cases often involve a large number of products or regions, where it may well be useful to study short- and long-run price relationships as a first step.

5. Conclusion

This paper considers the role of the law of one price in defining the relevant market for competition policy. We consider a range of conventional price tests used in market definition and also introduce a number of alternative and improved price tests. The performance of the conventional and new tests is then compared using data from a recent competition investigation in the South African milk industry. The main conclusions can be summarised as follows.

Firstly, the newer econometric procedures address some of the size and power problems faced by conventional tests. For example, based on the results in the empirical application, it is difficult to motivate the continued use (and therefore criticism) of the Dickey-Fuller unit root test when improved unit root tests are available. The newer tests are easy to apply and quite similar in structure to existing price tests, which recommends their use in practice.

Secondly, it is important to distinguish between price tests for short-run relationships and price tests for long-run relationships. Related, it is essential to understand that each price test asks a specific question when evaluating a relationship. As shown in the empirical application, results can differ quite significantly. For example, compare the results for the unit root tests and ARDL models. Unit root tests on price ratios detect stationarity between Western and Southern as well as Western and Eastern price pairs, but are inconclusive for the Southern and Eastern pair. However, the ARDL model finds a long-run equilibrium only for the Southern and Eastern price pair. Nonetheless, the unit root tests and ARDL models ask quite different questions: the former focuses on the existence of a one-to-one relationship, while the latter focuses on whether a *general* long-run relationship *exists and is significant* in the month-to-month behaviour of the prices. Therefore, a decision concerning the validity of a particular price test for market definition depends on the particular price relationship question that the practitioner considers important for market definition. For example, practitioners must decide on the relevant time horizon for market definition when deciding which price tests are useful to apply: if the practitioner believes short-run price relationships to be more relevant, he or she may prefer correlation analysis or Granger-causality tests. However, such a decision cannot be made a priori: 'long-run' adjustment can be as quick as six months or, alternatively, it could be relatively unimportant – implying a long run extending into a number of years. Therefore, it is advisable to run a batch of price tests that provide a rich perspective on price relationships, and then choose those tests which should be investigated further, based on the practitioner's view of the appropriate hypothesis. Multiple tests providing multiple answers should, therefore, be seen as a blessing rather than a curse.

Thirdly, the manner in which price tests are evaluated in the empirical application above illustrates that price tests ought to be used as *exploratory* tools for market definition, rather than as confirmatory tools. Price tests, as noted earlier, are based on a limited information set, and their results must be considered in the light of other descriptive or econometric evidence. In the empirical application, for example, the institutional change at the start of 2005 suggests that price tests must be sensitive to the inclusion or exclusion of 2005 data. More generally, econometric tools in market definition, and competition policy more generally, should be understood as exploratory tools. It is problematic to focus on finding a single encompassing test at the expense of the information offered by other types of tests. The noisiness of the data and the short length of the sample period are unlikely to support such an approach. This approach is consistent with the strategy adopted in most competition investigations, where evidence emerges in a piecemeal, evolutionary fashion. A variety of econometric tools assists in this discovery process. Where data challenges prevent more involved quantitative approaches, simpler tests such as modern price tests can form part of such a discovery process.

A proper understanding of price behaviour seems to us essential to understanding competition in any market and price tests appear to be useful in obtaining such an understanding – particularly where the more advanced econometric IO models are not feasible due to limited data. Of course, it would be very difficult to present any one price test statistic to a competition authority in support of a particular market definition. And it would be dangerous to rely on price tests only – but, as argued earlier in this paper, these are criticisms that one can also level at the more sophisticated econometric tools. However, a batch of price tests, properly executed can be a valuable empirical tool for competition analysts.

References

Abrantes-Metz, R.M. (2012) 'Why and how to use empirical screens in antitrust compliance?' *Competition Policy International Antitrust Chronicle*, February(1), pp. 1-8

Bishop, S., Walker, M. (1998) *Economics of E.C. Competition Law: Concepts, Application and Measurement*. Sweet and Maxwell, London.

Bishop, S., Walker, M. (2002) *The economics of EC competition law: concepts, application and measurement*, Second ed. Sweet and Maxwell, London.

Boshoff, W.H. (2007) 'Stationarity tests in geographic market definition: an application to South African milk markets', *South African Journal of Economics*, 75(1), pp. 1-14

Coe, P.J., Krause, D. (2008) 'An analysis of price-based tests of antitrust market delineation', *Journal of Competition Law and Economics*, 4(4), pp. 983-1007

Competition Tribunal (2008) 'Clover Industries Limited and Clover SA (Pty) Ltd versus the Competition Commission, Parmalat (Pty) Ltd, Ladismith Cheese (Pty) Ltd, Woodlands Dairy (Pty) Ltd, Lancewood (Pty) Ltd, Nestlé (Pty) Ltd and Milkwood (Pty) Ltd ', *78/CAC/Jul08*, South African Competition Tribunal.

Cournot, A. (1927) *Researches into the mathematical principles of the theory of wealth*. Macmillan, New York.

Davis, P., Garcés, E. (2010) *Quantitative techniques for competition and antitrust analysis*. Princeton University Press, Princeton.

Engle, R.F., Granger, C.W.J. (1987) 'Cointegration and Error Correction: Representation, Estimation and Testing', *Econometrica*, 55(2), pp. 251-276

Forni, M. (2004) 'Using stationarity tests in antitrust market definition', *American Law and Economics Review*, 6(2), pp. 441-464

Genesove, D. (2004) 'Comment on Forni's 'Using stationarity tests in antitrust market definition'', *American Law and Economics Review*, 6(2), pp. 476-478

Granger, C.W.J. (1969) 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica*, 37(pp. 424-438

Hadri, K. (2000) 'Testing for stationarity in heterogenous panel data', *Econometrics Journal*, 3(pp. 148-161

Haldrup, N. (2003) 'Empirical analysis of price data in the delineation of the relevant geographical market in competition analysis', *Working Paper 2003-09*, University of Aarhus.

Hosken, D., Taylor, C.T. (2004) 'Discussion of 'Using stationarity tests in antitrust market definition'', *American Law and Economics Review*, 6(2), pp. 465-475

Hunter, N.A. (2008) 'The market definition paradigm', *Department of economics*, University of California, Santa Barbara.

Im, K.S., Pesaran, M.H., Shin, Y. (2003) 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics*, 115(pp. 53-74

Katsoulacos, Y., Konstantakopoulou, I., Metsiou, E., Tsionas, E. (2012) 'Price comovement-tests in market definition: do they have a useful role?' *Competition and Regulation Summer School and Conference*, Crete, Greece.
Levin, A., Lin, C., Chu, C.J. (2002) 'Unit root tests in panel data: asymptotic and finite-sample properties', *Journal of Econometrics*, 108(pp. 1-24
Maddala, G.S., Kim, I.-M. (1998) *Unit roots, cointegration, and structural change*. Cambridge University Press, Cambridge UK.
Maphwanya, R., Muzata, G., Robb, G. (2012) 'The comforts and discomforts of using quantitative tests in defining markets with complex competitive dynamics: a review of evidence from a complaint and two mergers', *Sixth Annual Competition Commission, Competition Tribunal and Mandela Institute Conference on Competition Law, Economics and Policy*, University of the Witwatersrand, Johannesburg.
Marshall, A. (1920) *Principles of Economics*, Revised ed. Macmillan, London.
Massey, P. (2000) 'Market definition and market power in competition analysis: some practical issues', *The Economic and Social Review*, 31(4), pp. 309-328
McCarthy, T.R., Thomas, S.J. (2003) 'Geographic market issues in hospital mergers', National Economic Research Associates.
Mncube, L., Khumalo, J., Mokolo, R., Nijisane, Y. (2008) 'Use of price correlation and stationarity analysis in market definition - lessons from a recent merger', *Second Annual Competition Commission, Competition Tribunal and Mandela Institute Conference on Competition Law, Economics and Policy in South Africa*, University of the Witwatersrand, Johannesburg.
Narayan, P.K. (2005) 'The saving and investment nexus for China: evidence from cointegration tests', *Applied Economics*, 37(pp. 1970-1990
Ng, S., Perron, P. (2001) 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica*, 69(pp. 1519-1554
Pagan, A.R. (1989) '20 years after: Econometrics 1966-1986', in Cornet, B., Tulkens, H. (eds) *Contributions to Operations Research and Econometrics*, MIT Press, Cambridge, MA.
Pedroni, P. (2001) 'Purchasing power parity in cointegrated panels', *Review of Economics and Statistics*, 83(pp. 727-731
Perron, P. (1991) 'Test consistency with varying sampling frequency', *Econometric Theory*, 7(3), pp. 341-368
Pesaran, M.H., Shin, Y., Smith., R.J. (2001) 'Bounds Testing Approaches to the Analysis of Level Relationships', *Journal of Applied Econometrics*, 16(pp. 289-326
Stigler, G.J., Sherwin, R.A. (1985) 'The extent of the market', *Journal of Law and Economics*, 30(pp. 123-147
Turner, P. (2006) 'Response surfaces for an *F*-test for cointegration', *Applied Economics Letters*, 2006(13),

Appendix A: ARDL Bounds Test Methodology

Consider a bivariate $VAR(q)$ model for the price vector $= \begin{bmatrix} p_{1,t} \\ p_{2,t} \end{bmatrix}$. Assume all series are strictly $I(0)$, strictly $I(1)$ or cointegrated. The error term has zero conditional mean and is homoscedastic. The following *conditional* ECMs can be constructed from this VAR, the first conditioning on $p_{1,t}$ and the second on $p_{2,t}$:

$$\Delta p_{1,t} = \alpha_0 + \sum_{i=1}^{q} \alpha_{1,i} \Delta p_{1,t-i} + \sum_{i=0}^{q} \alpha_{2,i} \Delta p_{2,t-i} + \beta_1 p_{1,t-1} + \beta_2 p_{2,t-1} + \varepsilon_{p1,t}$$

$$\Delta p_{2,t} = \varphi_0 + \sum_{i=1}^{q} \varphi_{1,i} \Delta p_{2,t-i} + \sum_{i=0}^{q} \varphi_{2,i} \Delta p_{1,t-i} + \beta_3 p_{1,t-1} + \beta_4 p_{2,t-1} + \varepsilon_{p2,t}$$

From these models, one can test for the existence of a long-run relationship between $p_{1,t}$ and $p_{2,t}$, and if verified, estimate the long-run relationship. Formally, the structure of the null hypotheses and alternative hypotheses can be stated as follows:

$$H_0 = H_0^{\beta_1} \cap H_0^{\beta_2}$$

$$H_1 = H_1^{\beta_1} \cup H_1^{\beta_2}$$

where

$$H_0^{\beta_1}: \beta_1 = 0$$

$$H_0^{\beta_2}: \beta_2 = 0$$

$$H_1^{\beta_1}: \beta_1 \neq 0$$

$$H_1^{\beta_2}: \beta_2 \neq 0$$

The alternative hypothesis therefore also covers the degenerate cases $\beta_1 \neq 0$, $\beta_2 = 0$ and $\beta_1 = 0$, $\beta_2 \neq 0$. PSS show that, only under the assumption that $\beta_1 \neq 0$ can one derive a conditional level relationship for $p_{1,t}$ and $p_{2,t}$. If $\beta_2 = 0$, the conditional ECM clearly has no level effects, and there is no possibility of any level relationship. If $\beta_2 \neq 0$, then $\Delta p_{1,t}$ depends on the levels of $p_{2,t}$ *only* through its relation with the coefficients of $p_{2,t}$ in the original (not conditional) ECM, which does not point to a *long-run* relationship. As PSS acknowledge, there is still the potential for short-run relationships.

Null hypotheses of the above form involve multiple parameter restrictions and are usually tested using an F-statistic, which is compared with some critical value at a prescribed significance level. PSS compute asymptotic critical values for tests of this hypothesis on the conditional ECM of interest. The F-distribution for these joint hypotheses depends critically on the order of integration of the conditioning price variable. To avoid pretesting, PSS introduce a bounds testing approach relying on two critical values: simulation evidence suggests that the critical values for $I(0)$ conditioning variables form a lower bound, while critical values for the $I(1)$ case form an upper bound. To ensure that the conditional ECM is congruent with the underlying data, PSS develop separate critical values for different specifications involving deterministic components (trends and intercepts) of the VAR. Turner (2006) and Narayan (2005) develop corresponding small-sample critical values. Below are the small-sample critical bounds for sample sizes $n = 35$ and $n = 45$. These are closest to the actual sample sizes of between 36 and 48 in this case. For comparison, the asymptotic values originally proposed by PSS are also reported.

Table 6: Critical *F*-values for bounds test

Significance level	$\boldsymbol{n = 35}$		$\boldsymbol{n = 45}$		**Asymptotic**	
	$\boldsymbol{I(0)}$	$\boldsymbol{I(1)}$	$\boldsymbol{I(0)}$	$\boldsymbol{I(1)}$	$\boldsymbol{I(0)}$	$\boldsymbol{I(1)}$
1%	7.87	8.96	7.74	8.65	6.84	7.84
5%	5.29	6.18	5.24	6.14	4.94	5.73
10%	4.22	5.05	4.23	5.02	4.04	4.78

Appendix B: Panel Unit Root Test Results

Table 7: Panel unit root test results, January 2002 to December 2005

Lag	LLC	Breitung	IPS	ADF Fisher
1	-2.22** (0.01)	-1.42* (0.08)	-0.88 (0.19)	10.63 (0.10)
2	-1.09 (0.14)	-0.01 (0.49)	-0.16 (0.43)	7.73 (0.26)
3	-0.77 (0.22)	0.53 (0.70)	0.15 (0.56)	5.79 (0.45)
4	-0.66 (0.25)	0.49 (0.69)	-0.07 (0.47)	6.29 (0.39)
5	-0.25 (0.40)	0.33 (0.63)	0.65 (0.74)	2.60 (0.86)
6	0.58 (0.72)	0.51 (0.69)	1.11 (0.87)	1.38 (0.97)

Appendix C: Panel Cointegration Test Results

Table 8: Pedroni panel cointegration results

Lag	Dimension	January 2002 - December 2004	January 2002 – December 2005
1	Within	Cointegration	No cointegration
	Between	Cointegration	No cointegration
2	Within	Cointegration	No cointegration
	Between	Cointegration	No cointegration
3	Within	Cointegration	No cointegration
	Between	Cointegration	No cointegration
4	Within	Cointegration	No cointegration
	Between	No cointegration	No cointegration
5	Within	Inconclusive	No cointegration
	Between	No cointegration	No cointegration

Table 9: Johansen Fisher cointegration test results

Lag	Dimension	2002-2004	2002-2005
1	No cointegration	0.0038***	0.0004***
	One cointegration relation	0.0265**	0.0128**
2	No cointegration	**0.0012*****	0.0003***
	One cointegration relation	**0.2531**	0.0781*
3	No cointegration	**0.0329****	0.0271**
	One cointegration relation	**0.2505**	0.0521*
4	No cointegration	0.8656	0.6318
	One cointegration relation	0.4648	0.0798*
5	No cointegration	0.5662	0.0798*
	One cointegration relation	0.2935	0.1076
6	No cointegration	0.1107	0.1154
	One cointegration relation	0.2088	0.0292