AN EMPIRICAL ANALYSIS OF REAL COMMODITY PRICE TRENDS: AGGREGATION, MODEL SELECTION AND IMPLICATIONS

John T. Cuddington
Hong Wei
Georgetown University

Resumen: Este trabajo reexamina la validez empírica de la hipótesis de Presch-Singer acerca del deterioro secular de los términos de intercambio. Se contrastan los resultados empíricos obtenidos con índices aritméticos y geométricos. Cuando el índice aritmético de Gilli-Yang es usado, los resultados no son concluyentes. Un nuevo índice geométrico rinde, en contraste, resultados que son robustos.

Abstract: This paper re-examines the empirical validity of the Prebisch-Singer hypothesis of a secular decline in the relative price of primary commodities in terms of manufacturing goods. The empirical findings based on arithmetic and geometric commodity price indices are compared. When Grilli-Yang's arithmetic mean index is employed, the findings regarding commodity price trends are inconclusive. A new geometric mean index, in contrast, yields results that are robust.

1. Introduction and Summary of Findings

In the late 1940s, Raul Prebisch (1950) and Hans Singer (1950) independently argued that a number of factors would conspire to produce a secular deterioration in the relative price of primary commodities in terms of manufactured goods: *(i)* faster productivity growth in the manufacturing sector, *(ii)* asymmetric market structures, with competition in the primary commodity markets and oligopoly in the manufacturing sector, and *(iii)* higher income elasticities for manufacturing goods.[1] This negative relative price trend would, in turn, produce a deterioration in LDCs' net barter terms of trade (NBTT) to the extent that they continued to be net exporters of primary commodities and importers of manufactures.

The Prebisch-Singer (PS) hypothesis[2] has had a profound impact on economic policy making. During the 1960s and 1970s, it provided the justification for import-substitution policies of industrial development. The purpose of these policies, of course, was to diversify their productive structures and export bases away from primary commodities towards manufactured goods, thereby avoiding the terms of trade deterioration PS warned about.

Over the years, empirical analyses of commodity price trends have proliferated; few hypotheses in development economics have attracted as much attention as the PS hypothesis. The initial empirical investigations of Prebisch and the United Nations (1949) supported the theory. The evidence accumulated during the 1950s, 1960s, and 1970s, which has reflected gradual improvements in the quality of available commodity and manufacturing price indices and more sophisticated econometric techniques, has been mixed. (Recent literature is summarized in Section 2 below.)

To date, there has been little attention to the way in which commodity price indices have been constructed from the underlying commodities. This paper compares two alternatives: the arithmetic index, employed in most previous work, and a geometric index.[3] It shows that conclusions regarding the long-term trend in commodity prices are sensitive to the choice of index. Our comparison highlights the aggregation problems inherent in the use of aggregate indices (and sub-indices) when examining long-term price trends. Section 3 argues that the widespread use of arithmetic indices creates difficulties in relating trend movements in the index to those of its underlying commodity prices.[4] The geometric mean index is proposed as an alternative.

[1] There were, of course, countervailing views. The classical economists, for example, claimed that the NBTT of developing countries would improve over the long run because of diminishing returns in commodity production, increasing specialization and more rapid technical progress in manufacturing production, and growing population leading to increasing resource scarcity. See Grilli and Yang (1988) for a more detailed discussion of theoretical arguments regarding the PS hypothesis.

[2] The PS hypothesis is often taken to mean the secular deterioration in LDCs' terms of trade. In this paper, however, it refers to the hypothesized downward trend in the relative price of primary commodities in terms of manufactures, which is the supposed *cause* of the terms of trade deterioration.

[3] According to Spraos (1980), the Economist's commodity index was calculated as an arithmetic average until 1924, when a geometric index was adopted. We are aware of no other uses of the geometric index in tests of the PS hypothesis.

[4] See Cuddington (1990, forthcoming) for estimated time series models for the 24 commodities in the GY index plus oil and coal, each deflated by the MUV. The study finds that 16 of the 26 commodity prices are trendless. Five (five) have statistically significant negative (positive) trends.

Section 4 presents empirical results using the arithmetic index employed in previous work. In the process, the disparate findings of Grilli and Yang (1988) and Cuddington and Urzúa (1989), regarding the choice of *trend stationary* (TS) versus *difference stationary* (DS) models for this index, are reconciled. Although the unit root tests suggest that the TS model is appropriate for the arithmetic index, empirical results regarding the significance of a negative time trend are inconclusive. They depend critically on whether one believes that there is an exogenous break in the data in the early 1920s and accounts for it in the statistical analysis.

Section 5 presents the empirical results using our new geometric price index. With this index, standard unit root tests are somewhat inconclusive regarding the choice between the TS and DS specifications. Nevertheless, both models produce the same conclusions regarding the PS hypothesis: The trend in the geometric index of commodity prices (in terms of manufactures) index is insignificantly different from zero. Interestingly, this finding is upheld regardless of whether the 1921 data break is acknowledged or not.

2. A Selective Literature Review

Spraos (1980) was among those to rekindle academic interest in the empirical validity PS hypothesis in the late 1970s, at a time when enthusiasm for import substitution strategies in policy circles had faded. Following earlier studies, he estimated simple log-linear time trends for the NBTT using ordinary least squares regressions. Spraos concluded that for the 70-year period ending with the outbreak of World War II, there is indeed statistical evidence of a secular deterioration in the NBTT. Once the data sample period is extended to the post World War II period, however, the hypothesis is open to doubt.

In an extension of Spraos' work, Sapsford (1985) found evidence of discontinuous changes in the commodity price data used by Spraos. Taking the observed structural instability into account by allowing for intercept and slope dummies in 1950, Sapsford concluded that the evidence favored the PS hypothesis for the post World War II period as well as the pre-war period.[5]

The starting point for Grilli and Yang's (1988) work was their conviction, shared by Spraos, that having high quality price indices is essential when testing the PS hypothesis. They undertook a major data collection effort in order to construct a new commodity price index. The Grilli-Yang index, deflated by a manufacturing unit value (MUV) index, has subsequently been

[5] Unlike Spraos, Sapsford makes a much-needed Cochrane-Orcutt correction for first-order serial correlation.

used in a number of other studies (*e.g.* Cuddington and Urzúa, 1989; von Hagen, 1989; Cuddington, 1990 and Perron, 1990).

Using their new commodity index, Grilli and Yang re-estimated the time trend model and found a statistically significant long-term deterioration in the NBTT. Their econometric analysis paid attention to the possibility of first-order serial correlation as well as structural breaks. Regarding the latter, Grilli and Yang (1988, p. 10, our emphasis) note that, "examination of the residuals of the semilog time regressions, as well as a priori knowledge of the *exogenous* factors that may have caused a structural break in the price series, indicated the possibility of breaks at three points in time: 1921, 1932, and 1945."

Cuddington and Urzúa (1989) reexamined the Grilli-Yang index, paying special attention to the requirement that the error processes in the estimated time trend models be stationary (to ensure the validity of standard hypothesis tests). The possibilities of unit roots as well as structural breaks in the error process were considered. They concluded that there was no significant secular decline in the relative commodity prices once the statistical problems caused by non-stationary were addressed. When they adopted the TS specification and allowed for the large break in the data after 1920, the significant deterioration in commodity prices found by Grilli and Yang disappeared.

The TS model may not be appropriate however. The problem in applying a TS specification if the model is truly DS is that, although the coefficient estimates may be consistent,[6] the associated variance of the regression residuals are unbounded –even if there is, in fact, no trend. Thus, the standard hypothesis tests regarding the significance of the trend become invalid. On the basis of Said-Dickey-Perron tests for unit roots, Cuddington and Urzúa were unable to reject the hypothesis that the Grilli-Yang index has an unit root. Re-estimating the trend in commodity prices using the DS specification, however, they again found it to be trendless.

Perron (1990) has also examined the (MUV-deflated) Grilli-Yang price index (for 1900-1983) for the presence of a unit root. Using his "additive outlier" method for dealing with the structural break in 1920/1921,[7] which he argues is more appropriate when the change in the level of the time series is very sudden as is the case here, the unit root hypothesis is strongly rejected. Choosing the TS over the DS specification, he concluded that there was no significant trend term.

[6] Plosser and Schwert (1978) discuss conditions needed to insure that the OLS estimator is consistent in the presence of nonstationary regression residuals.

[7] This method involves a two-step procedure, detrending the data first with the appropriate dummies and then analyzing the remaining noise function for a unit root. Cuddington and Urzúa (1989), in contrast, use the method that Perron calls the innovational-outlier approach.

Von Hagen (1989) took a different approach from Cuddington and Urzúa (1989), but also stressed the time-series properties of the Grilli-Yang index (1900-1986). After finding that both the nominal commodity price and manufacturing unit values (MUV) indices were integrated of order one, he applied cointegration techniques in testing the PS hypothesis. The estimated cointegration vector between the (logarithms of the) nominal commodity price index and the MUV index was insignificantly different from unity. This suggested that the *relative* price has a constant long-run equilibrium value, not a secular trend. Therefore, von Hagen also concluded that his empirical evidence did not support the PS hypothesis.[8]

One troubling aspect of the Cuddington-Urzúa paper (as well as much of the other work just discussed) is their *ad hoc* methodology for isolating structural breaks. It was Cuddington and Urzúa's inspection of the residuals from simple TS and DS specifications that led them to conclude that there was a break in the Grilli-Yang index after 1920. This "data snooping" might be condemned because it in effect chooses a point that is likely to give a high Chow-test value. Strictly speaking, Perron's statistical test for unit roots in the presence of a shift in the mean require that the event be exogenous. Then it is valid to take it out of the stochastic process generating the data. Although Grilli and Yang (quoted above) allude to "exogenous" events that might explain the data break, they do not provide a discussion. Cuddington and Urzúa (1989, p. 429-430) admit that "in light of its critical importance in our findings below, this sharp drop in prices cries out for some explanation... Presumably, it reflects the adjustment in commodity supplies and demands following the end of the first World War." They go on to paraphrase Friedman and Schwartz's discussion of the belated tightening in Federal Reserve policy in 1920. While some case is made that the event is exogenous, a skeptic would not be convinced that this treatment of the break was appropriate. Unfortunately, the conclusion regarding the trend in commodity prices seems to hinge critically on whether or not the break is taken into account in the time series analysis.

3. Arithmetic versus Geometric Price Indices

This paper compares two commodity price indices, both constructed using Grilli and Yang's underlying price series on 24 non-fuel primary com-

[8] It is noteworthy that von Hagen includes *several* dummy variables in his analysis to capture structural breaks. Even if one takes these breaks to be exogenous, his Said-Dickey tests for unit roots are inappropriate in this situation. A generalization of Perron's test to the case of several data breaks would be required.

modities.[9] Each commodity price is deflated by their manufacturing unit value (MUV.) index series as in Grilli and Yang, so that *relative* commodity prices are considered throughout.[10] The data are annual and cover the period from 1900 to 1988. The first is the *arithmetic mean index* (AP_t) constructed by Grilli and Yang. It weights each (relative) commodity price p_{it} by its share of the total value of world commodity exports (a_i) in 1977-1979:

$$AP_t = \sum_{i=1}^{24} a_i \times p_{it} , \qquad (1)$$

where i and t refer to the commodity and time, respectively. The weights, of course, sum to unity.

The second index considered is a *geometric mean index* (GP_t), which we constructed using the same 24 commodity prices and weights utilized by Grilli and Yang:

$$GP_t = \prod_{i=1}^{24} p_{it}^{a_i} . \qquad (2)$$

As Figure 1 shows, however, the arithmetic and geometric indices are quite different. In light of this, one might ask: is there any conceptual reason to prefer one to the other? The arithmetic index has the advantage of a ready interpretation as the value of a fixed basket of goods (in terms of the MUV, i.e., units of industrial country exports). The geometric index is admittedly less intuitive, although it is the perfect price index for an economic agent with a Cobb-Douglas utility function over the 24 commodities.[11] More importantly, in the present context where the *log-linear* specification has been used in estimating long-term price trends, the logarithm of the geometric index is just the arithmetic average of the logarithms of the 24 underlying commodity prices. That is:

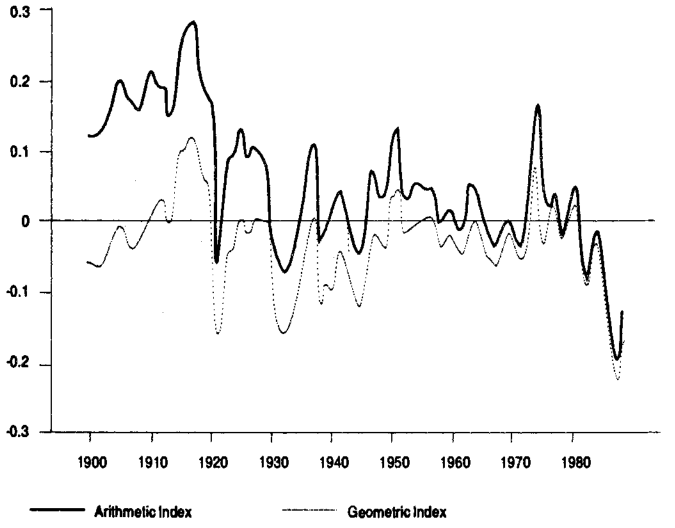
$$\ln(GP_t) = \sum_{i=1}^{24} a_i \times \ln(p_{it}). \qquad (3)$$

[9] These 24 commodities (listed in Table 1) cover 54% of all non-fuel commodities traded in the world in 1977-1979.

[10] Grilli and Yang obtained their MUV index from the United Nations' series by interpolating to fill in missing war-time observations. It reflects the U.S. dollar unit values of manufactured exports from a number of industrial market economies (France, Germany, Japan, United Kingdom and United States) to developing countries. Incidently, the MUV is an arithmetic index. The underlying data to recompute it as a geometric index are not readily available.

[11] Neither index, it should be acknowledged, gets particularly high marks from the perspective of modern index number theory (as they have constant weights over time and are, at best, linear approximations to more genenal utility functions, etc.).

Figure 1
Relative Price of Commodities



Suppose the geometric index has a time trend β_{GP} in the standard specification:

$$\ln(GP_t) = \alpha + \beta_{GP} \times \text{time} + e_t \,. \tag{4}$$

This time trend in the overall index should be equal to the weighted average of the time trends for the individual commodities in the index. Specifically, if the time trend for commodity i is β_i in:

$$\ln(p_{it}) = \alpha_i + \beta_i \times \text{time} + e_{it} \,, \tag{5}$$

then $\beta_{GP} = \sum_{i=1}^{24}(a_i \times \beta_i)$. The aggregation is natural.

The arithmetic index, in contrast, has a rather complicated, nonintuitive form after taking logarithms:

$$\ln(AP_t) = \ln(\sum_{i=1}^{24} a_i \times p_{it}) \,. \tag{6}$$

Because the logarithm operator is nondistributive, it is not easy to see how the trend in the arithmetic index, denoted β_{AP} say, is related to the trends in the underlying commodities (the β_i's in (5)). This relationship can be obtained, however, by applying the time-difference operator d to (6):

$$d\ln AP_t = \sum_{i=1}^{24} w_{it} \times d\ln p_{it} \tag{7}$$

where

$$w_{it} = a_i \times p_{it} / AP_t \ . \tag{8}$$

Noting that $d\ln AP_t / dt = \beta_{AP}$ on the left-hand side of (7) and that $d\ln p_t^i / dt = \beta_i$ from (5), the identity (7)-(8) can be rewritten as:

$$\beta_{AP} = \sum_{i=1}^{24} [w_{it} \times \beta_i + d\ln p_{it} \times dw_{it} / dt] \ . \tag{9}$$

Note that the trend in the arithmetic index in (9) can not be decomposed straightforwardly (i.e. linearly) in terms of the individual commodity price trends, because the implicit weights in (8) change over time. Even if all individual commodity prices are trendless ($\beta_i = 0$), the trend in the index, β_{AP}, may be nonzero. Thus, it is possible that a nonzero trend in the index is solely due to the process of aggregation. In general, implicit time-varying weights will make it difficult to interpret empirical findings regarding trends and cycles in arithmetic indices in terms of the behavior of the underlying commodity prices reported in Cuddington (1990, forthcoming).

As a practical matter, how important are the shifting implicit weights w_{it} in the AP index? Do they differ significantly from the trade weights a_i specified by Grilli and Yang? As Table 1 shows, the differences between a_i and w_{it} can be very large indeed. Coffee, for example, had a trade weight of 10.3 percent in 1977-1979. In contrast, the implicit weight (w_{it}) in the arithmetic index averages only 4.8 percent, but ranges from a low of 1.5 to a high of 13.8 percent; see Figure 2.

4. Unit Roots and the Trend in the Grilli-Yang Index

As Section 2 highlights, the empirical conclusions regarding the trend in the Grilli-Yang (arithmetic) index hinge critically on two issues, the presence of unit roots and structural breaks. Cuddington and Urzúa (1989) considered the general ARIMA specification:

$$ly(t) = a + \beta \times \text{time} + b \times D1921 + e(t) \tag{10}$$

$$(1 - \rho L)A(L)e(t) = B(L)u(t) \ . \tag{11}$$

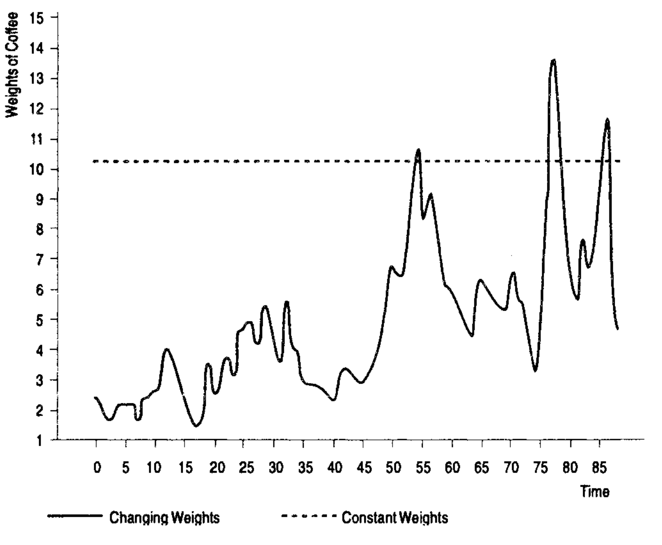
where $ly(t)$ is the natural log of the chosen price index and β is the time trend. $D1921$ is a dummy variable that takes the value of unity from 1900 to 1920, and zero thereafter. The error process in (11) has been factored to isolate the root with the largest modulus ρ. It is assumed that: (*i*) the lag polynomial $A(L)$ is invertible, (*ii*) $B(L)$ and hence $A^{-1}(L) \times B(L)$ are stable polynomial lag operators, and (*iii*) the innovations $u(t)$ in (11) are white noise.

Table 1
Comparison of the Implicit Time-Varying Weights with the Given Constant Weights

Primary Commodity	*Grilli-Yang Weights* (a_i)	*Implicit Weights* (W_{it}) *Average*	*Difference* [1]	*Max.*	*Min.*
Coffee	10.3	4.8	5.5	13.8	1.5
Cocoa	2.7	1.0	1.7	3.1	0.4
Tea	1.6	1.7	-0.1	2.9	0.7
Wheat	8.1	10.6	-2.5	16.7	7.1
Maize	6.8	9.2	-2.4	15.1	5.5
Rice	3.0	3.6	-0.6	5.4	1.7
Sugar	7.3	9.4	-2.1	25.0	3.8
Beef	5.1	2.4	2.7	6.5	0.6
Lamb	0.9	0.4	0.5	1.1	0.1
Palm	8.3	8.5	-0.2	19.2	4.1
Banana	0.9	1.1	-0.2	2.3	0.5
Cotton	4.3	5.3	-1.0	8.6	3.0
Wool	2.7	4.8	-2.1	7.8	2.1
Rubber	2.8	6.6	-3.8	24.4	1.7
Timber	12.0	7.0	5.0	12.2	2.5
Jute	0.2	0.2	0.0	0.5	0.1
Tobacco	2.9	2.3	0.6	4.4	0.6
Hide	2.3	2.3	0.0	4.0	1.0
Cooper	5.9	5.6	0.3	10.8	3.3
Aluminum	5.1	8.7	-3.6	20.2	3.1
Silver	1.7	0.8	0.9	3.6	0.4
Tin	2.2	1.0	1.2	2.4	0.3
Lead	1.3	1.0	0.3	1.6	0.5
Zinc	1.6	1.5	0.1	2.5	0.8
TOTAL	100.0	100.0			

[1] The difference is equal to the Grilli-Yang weight minus the average of the implicit weights.

Figure 2
Changing versus Constant Weights for Coffee



As long as $|\rho| < 1$ (10)-(11) is just a general ARMA version of the TS model on which most previous analyses of the PS hypothesis are based. If there is a unit root ($\rho = 1$), (10)-(11) becomes the DS model. Because the error process is nonstationary in this case, however, the sampling distribution of the OLS estimator of β in (10) is not well behaved. This may lead to OLS estimates that are not only biased but, in many applications, also inconsistent.[12] In fact, it is extremely likely that researchers would conclude that commodity prices exhibit trends even if no trends exist. (Nelson and Kang, 1984).

In the presence of a unit root, the appropriate procedure is to first-difference (10)-(11) to eliminate the unit root and achieve stationarity of the error process. The estimating equation becomes:

$$dly(t) = \beta + n(t) \tag{12}$$

where

$$A(L)n(t) = B(L)u(t) \; . \tag{13}$$

The coefficient β captures the (stochastic) time trend in $ly(t)$.

[12] Although the OLS estimator is always biased, Plosser and Schwert (1978) show that it turns out to be consistent in the special case where TIME is the only regressor (as it is in many of the empirical studies of commodity price behavior).

Just as under-differencing (i.e. failing to recognize the presence of an unit root) can lead to incorrect statistical inferences, so can over-differencing. If there is no unit root in (11), it is inappropriate to first-difference the model prior to estimation. Doing so will *introduce* an unit root into the MA part of the error process of the first-differenced form of the model in (12)-(13) if none existed in the original specification (10)-(11). Thus, it is clearly desirable to *test* for the presence of an unit root prior to estimation of the time trend.

Cuddington and Urzúa's analysis uses the unit root tests suggested by Said-Dickey (1984) and extended by Perron (1989) for time series with structural breaks.[13] The Said-Dickey-Perron test is based on the regression equation:

$$d\ln y_t = \rho \times \ln y_{t-1} + \alpha_0 + \alpha_1 \times \text{time} + \alpha_2 \times D1921$$
$$+ \alpha_3 \times DD1921 + \beta(L) \times d\ln y_{t-1} + e_t \qquad (14)$$

where the y_t is the chosen (relative) price index. $D1921$ is the level-shift dummy described above and $DD1921$ is a "spike" dummy that takes the value zero in every year except 1921 when it equals one. The null hypothesis that there exists an unit root, implies that $\rho = 0$ in equation (14). If ρ turns out to be significantly different from zero then the existence of unit root hypothesis is rejected.

When implementing the Said-Dickey or Perron tests, an important consideration is the choice of lag length for the polynomial $\beta(L)$. If an excessive number of lags is used, the unit root tests lack power. On the other hand, if the lag length is inadequate, the residuals in (14) will not be white noise and the t-statistics will not have the usual $N(0, 1)$ distribution.

Table 2 reports the Perron test results for the GY index.[14] Per Perron's (1990) suggestion, the additive outlier method is used for incorporating the data break into the unit root tests. Alternative lag lengths from zero to five for $\beta(L)$ are considered.[15] As the Table shows, the conclusion regarding the

[13] Perron shows that the Said-Dickey tests are invalid if an exogenous structural break is present. Given the apparent break in the data in 1921, Cuddington and Urzúa chose the Perron test. It is based on (14). The Said-Dickey test is based on a similar regression except that the dummies $D1921$ and $DD1921$ are omitted, and the distribution of the key parameter ρ is different.

[14] Readers who are skeptical about the presence of an "exogenous" data break in 1920/1921 may prefer to look at the Said-Dickey unit root tests in Appendix 1 instead.

[15] Given that the conclusions from the unit root tests are often quite sensitive to the choice of lag length in (14), it is desirable to have some criteria for selecting the appropriate lag-length in (14). Two frequently employed criteria are the Akaike information (AIC) and Schwarz (SC) criteria. The values of these criteria for each model are shown in the rightmost columns of Table 2; smaller values indicate preferred models. The model that would be chosen by Perron's (1989, p. 1382), which is based on the size of the t-values on the coefficients on the various lags of the dependent variable, is also indicated.

presence of a unit root (which is based on Perron's distribution for t-value for ρ) is insensitive to the chosen lag length.[16] The unit root hypothesis is rejected. This result provides support for choosing the TS model, as Grilli and Yang did when estimating the trend in their arithmetic index.[17]

Table 2
Perron Unit Root Tests for Arithmetic Index (1900-1988)
(Additive Outlier Method)

$$\ln(AP_t) = \alpha + \beta \times \text{time} + \beta_1 \times D1921 + y_t$$
$$dy_t = \rho \times y_{t-1} + \alpha + \beta_3(L) \times y_{t-1} + e_t$$

Model	*t-Stat. on* ρ	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-3.16	83	10	.0105	.0123	
AR4	-3.51*	84	9	.0103	.0117	
AR3	-3.66*	85	8	.0101	.0112	
AR2	-3.85**	86	7	.0098	.0107	
AR1	-4.82**	87	6	.0096	.0102	###
AR0	-4.00**	88	5	.0091	.0093	

Notes: * Significant at 10% level. Reject the unit root hypothesis. ** Significant at 5% level. Reject the unit root hypothesis.

The critical values for 5% and 10% significant levels with the structural break point at $\lambda = 21/88$ are -3.77 and -3.47.

The t-Stat. is the t-Statistic of the coefficient p in the regression. T and K denote the number of observations and parameters, respectively. AIC and SC are the values of Akaike information and Schwarz criteria. The lag length that would be chosen using Perron's (1989, p. 1382) method is indicated by ###.

Having rejected the unit root hypothesis, the TS model is used to estimate the time trend in the (deflated) GY arithmetic index. The results are reported in Table 3. Regression (15) essentially replicates the findings of Grilli and Yang. It shows that, ignoring any break in the data after 1920, the index exhibits a statistically significant negative trend; the errors follow an AR(1) process. This

[16] When Perron's so-called innovational outlier method is used, as was done in Cuddington and Urzúa (1989), similar results are obtained. See Appendix 3 (available upon request; not for publication, but attached for referees' information).

[17] Cuddington and Urzúa (1989) used the Said-Dickey-Perron test for the presence of a unit root in the Grilli-Yang index using 1900-1983 data. They considered only lag lengths from 4 to 10 for L in (14), employing the innovational outlier method. Unlike Grilli and Yang, they found that the unit root hypothesis could not be rejected. This led them to conclude that the DS model is probably more appropriate than the TS model when estimating the trend in the GY index, although they obtain consistent conclusions from the two specifications.

led Grilli and Yang to their positive conclusion regarding the validity of the PS hypothesis.

Table 3
Trend Stationary Models for the Grilli-Yang Index (1900-1988)

$$\ln AP_t = .402 - .007 \times \text{time} + e_t \quad (15)$$
$$\quad (4.691) \quad (-4.035)$$

$$e_t = .720 \times e_{t-1} + \varepsilon_t \qquad SER = .107$$
$$(9.504)$$

$$\ln AP_t = .111 - .002 \times \text{time} + .382 \times D1921 + e_t \quad (16)$$
$$(1.481) \quad (-1.372) \quad (5.273)$$

$$e_t = .881 \times e_{t-1} - .271 \times e_{t-2} + \varepsilon_t \qquad SER = .093$$
$$(8.204) \quad (-2.396)$$

$$\ln AP_t = .176 - .003 \times \text{time} + .298 \times D1921 + .153 \times DD1921 + e_t \quad (17)$$
$$(2.257) \quad (-2.091) \quad (3.738) \quad (1.942)$$

$$e_t = .860 \times e_{t-1} - .268 \times e_{t-2} + e_t \qquad SER = .091$$
$$(8.002) \quad (-2.358)$$

Notes: 1) The dummy *D*1921 takes the value of unity from 1900 through 1920; it is zero thereafter.
2) The dummy *DD*1921 takes the value of zero in every year except 1921 when it equals one.
3) *SER* stands for standard error of regression.
4) The numbers in parentheses are *t* values.

In light of the apparent downward shift in the (deflated) GY index after 1920, Cuddington and Urzúa included the level-shift dummy *D*1921 in their time series models. The result in regression (16) shows that, with the inclusion of the highly significant dummy, the coefficient on time is no longer significant.[18]

In the process of inspecting the regression residuals in (16), a large negative residual (i.e. more than three standard deviations from zero) in the year 1921 was detected. To examine the importance of this outlier, a spike dummy *DD*1921 was added to regression (16); the result is (17). The level-shift dummy *D*1921 remains significant. The new spike dummy *DD*1921

[18] These results are similar to those earlier obtained by Cuddington and Urzúa (1989), although their sample period ranges from 1900-1983 rather than 1900-1988.

is significant at the 10 percent, although not at the 5 percent level. Interestingly, a statistically significant negative trend reappears!

In sum, the empirical results using Grilli and Yang's arithmetic index are somewhat inconclusive. They depend on whether and how one accounts for the structural break after 1920.

5. Estimating Trends with Geometric Price Index

The geometric index can now be analyzed using the methodology employed for the arithmetic index in Section 4. Because the break in the geometric index in 1921 is rather minor, Table 4 reports the results of the Said-Dickey unit root test for alternative lag specifications of $\beta(L)$ in (14). (The conclusions below are unchanged if the Perron test is used, as one would expect in this situation; see Appendix 1.)

Table 4

Said-Dickey Unit Root Tests for Geometric Index

$$d\ln GP_t = \rho \times \ln Gp_{t-1} + \alpha + \beta \times \text{time} + \beta_1(L) \times d\ln GP_{t-1} + e_t$$

Model	*t-Stat*	T	K	*AIC*	*SC*	*Perron*
AR5	-2.51	83	8	.0103	.0125	
AR4	-2.78	84	7	.0100	.0118	
AR3	-2.75	85	6	.0091	.0101	
AR2	-2.87	86	5	.0093	.0106	
AR1	-3.81**	87	4	.0091	.0101	###
AR0	-3.26**	88	3	.0092	.0100	

Notes: * Significant at 10% level. ** Significant at 5% level.

The critical values, computed using the response function in Mackinnon (1990), for 5% and 10% significant levels of 88 sample observations are -3.46 and -3.15, respectively. The lag length that would be chosen using Perron's (1989, p. 1382) method is indicated by ###.

T and K denote the number of observations and parameters, respectively. The t-Stat. is the t-Statistic of the coefficient ρ in the regression. AIC and SC are the values of Akaike information and Schwarz criteria.

Attempting to use the Akaike information, Schwarz, and Perron criteria to select the appropriate lag specification gives rise to no clearcut choice. Yet, the t-statistics on ρ yield conflicting conclusions about the unit root hypothesis. In short, the Said-Dickey test (or the Perron test in Appendix 1) does not enable us to choose between the TS and DS models. For this reason, the trend in the index is estimated using the TS and DS models in turn. Table 5A shows several TS specifications in order to facilitate a comparison with Table 3 above. Most strikingly, the TS models for the geometric index show no significant time trend, regardless of whether the level-shift dummy *D*1921 is included. The spike dummy *DD*1921 is insignificant and –unlike regression (17) above– its inclusion does not alter our conclusions regarding the trend in commodity prices.

Table 5B shows the DS models. The trend in real commodity price index, reflected in the constant terms in the DS regressions, is insignificantly different from zero, as it was in the TS specifications. Thus, regardless of whether one believes there is/is not unit root in the statistical process for the geometric index (because of the inconclusive unit root tests in Table 4), none of the regressions in Table 5 show a statistically significant trend in commodity prices over the 1900-1988 period.[19]

Comparing the results in Section 4 and 5 leads to the following conclusions: arithmetic index produces rather clear cut results rejecting the unit root hypothesis (Table 2), but the conclusions regarding the trend in the index are very sensitive to minor changes in model specification (Table 3). Thus it is difficult to reach firm conclusions regarding the PS hypothesis.

For the geometric price index, on the other hand, the unit root tests were inconclusive (Table 4). Nevertheless, the conclusions regarding the trend in commodity prices remain the same regardless of whether *(i)* the TS or DS specification is chosen and *(ii)* a structural break in the data after 1920 is entertained: The trend is found to be statistically insignificant. Thus these findings add to the weight of recent evidence against the empirical validity of the Prebisch-Singer hypothesis.

[19] An interesting alternative test for the presence of a significant time trend is suggested by the work of Pierre Perron: Perform the unit root test *without a time trend term*. If a rejection is possible, this gives indirect evidence that the trend term is insignificant. The logic is as follows. Perron (1988, Theorem 1, p. 316) argues that if the data is generated by a TS process with a nonzero trend, the unit root test that omits the trend term will fail to reject the unit root hypothesis even asymptotically. The intuition for this result is well explained by Perron (1988, p. 316): "Suppose a series shows a definite tendency to have an increase in its mean over time and a regression is used with only a constant and a lagged dependent variable. The only way to capture this increase is to make the constant become a drift term which occurs when the autoregressive parameter is set at one." (Though this result was derived for the case of no breaks, the same conclusion applies.) Following this argument, unit root tests excluding trend term were run for arithmetic and geometric indices. The results, which are shown in Appendix 2, are sensitive to the choice of lag length. Interestingly, the unit root hypothesis is rejected for the AR1 model, which is selected using any of the three (Akaike information, Schwarz, or Perron) criteria. Thus, these results provide indirect evidence that the trend term is insignificant. Perron (1990) reached the same conclusion for the arithmetic index (with the additive outlier method) over the shorter sample 1900-1983.

Table 5A

Trend Stationary Models for the Geometric Index (1900-1988)

$$\ln GP_t = -.009 - .002 \times \text{time} + e_t \qquad (18)$$
(.009) (-1.204)

$$e_t = .759 \times e_{t-1} + \varepsilon_t \qquad SER = .094$$
(10.307)

$$\ln GP_t = -.127 - .000 \times \text{time} + .386 \times D1921 + \varepsilon_t \qquad (19)$$
(-.0541) (-.112) (4.368)

$$e_t = .905 \times e_{t-1} - .252 \times \varepsilon_{t-2} + \varepsilon_t \qquad SER = .086$$
(9.302) (-2.264)

$$\ln GP_t = -.116 - .001 \times \text{time} + .261 \times D1921 - .121 \times DD1921 + \varepsilon_t \qquad (20)$$
(-.629) (-.111) (2.237) (-1.380)

$$e_t = .863 \times e_{t-1} - .216 \times \varepsilon_{t-2} + \varepsilon_t \qquad SER = .085$$
(9.686) (-1.563)

Table 5B

Difference Stationary Models for the Geometric Index (1900-1988)

$$d\ln GP_t = -.003 + e_t \qquad (21)$$
(-.288)

$$e_t = \varepsilon_t - .283 \times \varepsilon_{t-2} \qquad SER = .096$$
(-2.627)

$$d\ln GP_t = .002 - .401 \times D1921 + e_t \qquad (22)$$
(.169) (-4.532)

$$e_t = \varepsilon_t - .248 \times e_{t-2} \qquad SER = .088$$
(-2.258)

Notes: 1) The dummy *D*1921 takes the value of unity from 1900 through 1920; it is zero thereafter.

2) The dummy *DD*1921 is the first difference of *D*1921. It takes the value of zero in every year except 1921 when it equals one.

3) *SER* stands for standard error of regression.

4) The numbers in parentheses are t values.

Appendix 1

Said-Dickey Unit Root Tests for Arithmetic Index

$$d\ln AP_t = \rho \times \ln AP_{t-1} + \alpha + \beta \times \text{time} + \beta_1(L) \times d\ln AP_{t-1} + e_t$$

Model	*t-Stat*	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-2.61	83	8	.0135	.0164	
AR4	-2.79	84	7	.0130	.0155	
AR3	-3.04	85	6	.0118	.0131	
AR2	-3.19*	86	5	.0122	.0138	
AR1	-3.98**	87	4	.0118	.0138	###
AR0	-3.69**	88	3	.0117	.0127	

Perron Unit Root Tests for Geometric Index
(Additive Outlier Method)

$$d\ln GP_t = \rho \times \ln GP_{t-1} + \alpha + \beta \times \text{time} + \beta_1 \times D1921 + \beta_2 \times DD1921 + \beta_3(L) \times d\ln GP_{t-1} + e$$

Model	*t-Stat.*	*T*	K	*AIC*	*SC*	*Perron*
AR5	-2.60	83	10	.0090	.0105	
AR4	-2.97	84	9	.0088	.0100	
AR3	-2.83	85	8	.0087	.0097	
AR2	-2.97	86	7	.0085	.0092	
AR1	-4.01**	87	6	.0084	.0089	###
AR0	-3.25*	88	5	.0078	.0080	

Note: * Significant at 10% level. ** Significant at 5% level.

Perron's critical values for 5% and 10% significant levels with the structural break at $\lambda = 21/88$ are -3.77 and -3.47. For the Said-Dickey test, the critical values, computed using the response function in MacKinnon (1990) for 5% and 10 % significant levels of 88 sample observations are -3.47 and -3.16, respectively. The lag length that would be chosen using Perron's (1989, p. 1382) method is indicated by ###.

T and K denote the number of observations and parameters, respectively. The *t*-Stat. is the *t*-Statistic of the coefficient ρ in the regression. AIC and SC are the values of Akaike information and Schwarz criteria.

Appendix 2

Perron Unit Root Tests for Arithmetic Index
(Additive Outlier Method, without Trend Term)

$$d\ln AP_t = \rho \times \ln AP_{t-1} + \alpha + \beta_1 \times D1921 + \beta_1(L) \times d\ln AP_{t-1} + e_t$$

Model	*t-Stat.*	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-2.34	83	8	.0100	.0116	
AR4	-2.77	84	7	.0096	.0109	
AR3	-2.88	85	6	.0093	.0103	
AR2	-3.18	86	5	.0090	.0097	
AR1	-4.25**	87	4	.0087	.0092	###
AR0	-3.41	88	3	.0091	.0093	

Perron Unit Root Tests for Geometric Index
(Additive Outlier Method, without Trend Term)

$$d\ln GP_t = \rho \times \ln GP_{t-1} + \alpha + \beta_1 \times D1921 + \beta_1(L) \times d\ln GP_{t-1} + e_t$$

Model	*t-Stat.*	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-2.72	83	10	.0087	.0101	
AR4	-3.07	84	9	.0084	.0095	
AR3	-3.00	85	8	.0082	.0091	
AR2	-3.11	86	7	.0079	.0086	
AR1	-4.09**	87	6	.0077	.0082	###
AR0	-3.39	88	5	.0079	.0082	

Note: * Significant at 10% level. ** Significant at 5% level.

Perron's critical values for 5% and 10% significant levels with the structural break at $\lambda = 21/88$ are -3.77 and -3.47. The lag length that would be chosen using Perron's (1989, p. 1382) method is indicated by ###.

T and K denote the number of observations the parameters, respectively. The *t*-Stat. is the *t*-Statistic of the coefficient ρ in the regression. AIC and SC are the values of Akaike information and Schwarz criteria.

Appendix 3

Perron unit Root Tests Using the Innovational Outlier Method
Arithmetic Index (Grilli-Yang)

$$d\ln AP_t = \rho \times \ln AP_{t-1} + \alpha + \beta \times \text{time} + \beta_1 \times D1921 + \beta_2(L) \times d\ln AP_{t-1} + e_t$$

Model	*t-Stat.*	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-4.08**	83	8	.0122	.0151	
AR4	-4.30**	84	7	.0118	.0143	
AR3	-4.54**	85	6	.0107	.0121	
AR2	-4.62**	86	5	.0110	.0128	
AR1	-5.33**	87	4	.0107	.0121	###
AR0	-4.64**	88	3	.0111	.0123	

Geometric Index (Cuddington-Wei)

$$d\ln GP_t = \rho \times \ln GP_{t-1} + \alpha + \beta \times \text{time} + \beta_1 \times D1921 + \beta_2 \times d\ln GP_{t-1} + e_t$$

Model	*t-Stat.*	*T*	*K*	*AIC*	*SC*	*Perron*
AR5	-3.06	83	10	.0102	.0126	
AR4	-3.32	84	9	.0098	.0119	
AR3	-3.31	85	8	.0090	.0102	
AR2	-3.40	86	7	.0092	.0107	
AR1	-4.30**	87	6	.0090	.0102	###
AR0	-3.59*	88	5	.0092	.0102	

Note: * Significant at 10% level. ** Significant at 5% level.

Perron's critical values for 5% and 10% significant levels with the structural break at $\lambda = 21/88$ are -3.77 and -3.47. The lag length that would be chosen using Perron's (1989, p. 1382) method is indicated by ###.

T and K denote the number of observations the parameters, respectively. The *t*-Stat. is the *t*-Statistic of the coefficient ρ in the regression. AIC and SC are the values of Akaike information and Schwarz criteria.

References

Box, G. E. P., and G. M. Jenkins (1970). *Time Series Analysis, Forecasting and Control*, San Francisco, Holden Day.

Cuddington, J. T. (1990). "Long-Run Trends in 26 Primary Commodity Prices: A Disaggregated Look at the Prebisch-Singer Hypothesis", *Journal of Development Economics*, forthcoming.

———, and C. M. Urzúa (1989). "Trends and Cycles in the Net Barter Terms of Trade: A New Approach", *Economic Journal*, June, pp. 426-442.

Dickey, D. A., W. R. Bell, and R. B. Miller (1986). "Unit Roots in Time Series Models: Tests and Implications", *American Statistician*, February, pp. 12-26

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, New York, Wiley.

Grilli, E. R., and M. C. Yang (1988). "Primary Commodity Prices, Manufactured Goods Prices, and Terms of Trade of Developing Countries: What the Long Run Shows", *World Bank Economic Review*, January, pp. 1-48.

MacKinnon, J. G. (1990). "Critical Values for Cointegration Tests", Queen's University and University of California, San Diego, manuscript.

Nelson, C. R., and H. Kang (1984). "Pitfalls in the Use of Time as an Explanatory Variable in Regression", *Journal of Business and Economic Statistics*, pp. 73-82.

Perron, P. (1988). "Trend and Random Walks in Macroeconomic Time Series, Further Evidence from a New Approach", *Journal of Economic Dynamics and Control*, vol. 12, pp. 297-332.

——— (1989). "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica*, November, pp. 1361-1401.

——— (1990). "Testing for a Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics*.

Plosser, C. I., and G. W. Schwert (1978). "Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing", *Journal of Monetary Economics*, vol. 4, pp. 637-660.

Prebisch, R. (1950). *The Economic Development of Latin America and its Principal Problems*, United Nations, Lake Success.

Said, S. E., and D. A. Dickey (1984). "Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order", *Biometrika*, pp. 599-607.

Sapsford, D. (1985a). "The Statistical Debate on the Net Barter Terms of Trade between Primary Commodities and Manufactures: A Comment and some Additional Evidence", *Economic Journal*, pp. 781-788.

——— (1985b). "Real Primary Commodities Prices: An Analysis of Long-Run Movements", Document DM/85/31, International Monetary Fund.

Scandizzo, P. L., and D. Diakosawas (1987). *Instability of the Terms of Trade of Primary Commodities: 1900-1982*, FAO Economic and Social Development Paper, vol. 64, Rome, FAO.

Singer, H. (1950). "The Distributions of Gains between Investing and Borrowing Countries", *American Economic Review, Papers and Proceedings*, May, pp. 473-485.

Spraos, J. (1980). "The Statistical Debate on the Net Barter Terms of Trade between Primary Commodities and Manufactures", *Economic Journal*, March, pp. 107-128.

Thirlwall, A. P. and Bergevin, J. (1985). "Trends, Cycles and Asymmetries in the Terms of Trade of Primary Commodities from Developed and Less Developed Countries", *World Development*, July, pp. 805-817.

United Nations (1949). *Relative Prices of Exports and Imports of Under-Developed Countries*, New York, Lake Success.

von Hagen, J. (1989). "The Relative Commodity Prices and Cointegration", *Journal of Business and Economic Statistics*, October, pp. 497-503.
World Bank (1982). *Price Prospects for Major Primary Commodities*, Report No. 8141-2, Washington.