No. 2002-06-A

OFFICE OF ECONOMICS WORKING PAPER
U.S. INTERNATIONAL TRADE COMMISSION

Prebisch-Singer Redux

John T. Cuddington
Georgetown University

Rodney Ludema
U.S. International Trade Commission
and Georgetown University

and

Shamila A Jayasuriya
Georgetown University

June 2002

Address correspondence to:
Office of Economics
U.S. International Trade Commission
Washington, DC 20436 USA

Prebisch-Singer Redux

John T. Cuddington, Rodney Ludema, and Shamila A Jayasuriya
Georgetown University

January 2002 draft

Abstract:

In light of ongoing concern about commodity specialization in Latin America, this paper revisits the argument of Prebisch (1950) that, over the long term, declining terms of trade would frustrate the development goals of the region. This paper has two main objectives. The first is to clarify the issues raised by Prebisch and Singer (1950), as they relate the commodity specialization of developing countries (and Latin America in particular). The second is to reconsider empirically the issue of trends in commodity prices, using recent data and techniques. We show that rather than a downward trend, real primary prices over the last century have experienced one or more abrupt shifts, or "structural breaks," downwards. The preponderance evidence points to a single break in 1921, with *no* trend, positive or negative, before or since.

1. MOTIVATION

Development economists have long debated whether developing countries should be as specialized as they are in the production and export of primary commodities. Nowhere has this question been debated morehotly than in Latin America. Indeed, it was Latin America that provided the motivation for the seminal contribution of Prebisch (1950) on this topic. He, along with Singer (1950), argued that specialization in primary commodities, combined with a relatively slow rate of technical progress in the primary sector and an adverse trend in the commodity terms of trade, had caused developing economies to lag behind the industrialized world. Prebisch concluded that, "since prices do not keep pace with productivity, industrialization is the only means by which the Latin-American countries may fully obtain the advantages of technical progress." Debate over the validity of Prebisch and Singer's claims, as well as the appropriate policy response, has occupied the literature ever since.

While much has happened in Latin America since 1950, the concern about specialization remains as topical as ever. According to noted economic historian and political economist Rosemary Thorp of Oxford University, "The 1990s already saw a return to a primary-exporting role for Latin America. All the signals are that the world economy will push Latin America even more strongly in this direction in the new century, especially in the fields of oil and mining. It behooves us to look very coldly at the political economy and social dimensions of such a model, with more than half an eye on the past. We need to be alert to what will need to change if primary-resource-based growth is to be compatible with long-term economic and social development."[1]

In light of this ongoing concern about commodity specialization in Latin America, we believe it is important to revisit Prebisch's concern of over 50 years ago that, over the long term, declining terms of trade would frustrate the development goals of the region. This paper has two main objectives. The first is to clarify the issues raised by Prebisch and Singer, as they relate the commodity specialization of developing countries (and Latin America in particular). The second is to reconsider empirically the issue of trends in commodity prices, using recent data and techniques.

2. THE PREBISCH-SINGER HYPOTHESIS

The Prebisch-Singer hypothesis normally refers to the claim that the relative price of primary commodities in terms of manufactures shows a downward trend. However, as noted earlier, Prebisch and Singer were concerned about the more general issue of a rising per capita income gap between industrialized and developing countries and its relationship to international trade. They argued that international specialization along the lines of "static" comparative advantage had excluded developing countries from the fruits of technical progress that had so enriched the industrialized world.

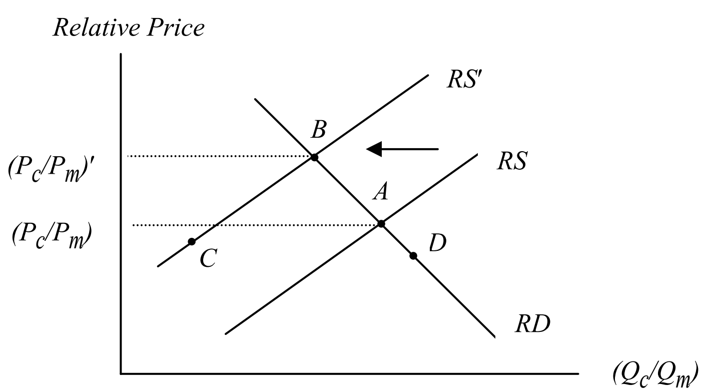
They rested their case on three stylized facts: first, that developing countries were indeed highly specialized in the production and export of primary commodities; second, that technical progress was concentrated mainly in industry; and third, that the relative price of primary commodities in terms of manufactures had fallen steadily since the late 19th Century. Together these facts suggested that, because of their specialization in primary commodities, developing countries had obtained little benefit from industrial

[1]Abstract of a lecture given at the Inter-American Development Bank on August 1, 2001.

technical progress, either directly, through higher productivity, or indirectly, through improved terms of trade.[2]

To see this point more clearly, consider Diagram 1, which offers a simple model of the world market for two goods, primary commodities and manufactures. The vertical axis measures the relative price of primary commodities in terms of manufactures, or P_c / P_m, while the horizontal axis measures relative quantities, the total quantity of commodities sold on the world market divided by the total quantity manufactures. The intersection of the relative demand (*RD*) and relative supply (*RS*) schedules determines the world market equilibrium.

Diagram 1: World Market for Primary Commodities Relative to Manufactures



If technical progress in the manufacturing sector exceeds that of the primary sector (as Prebisch and Singer supposed), then we should see the supply of manufactures growing faster than the supply of commodities. This would correspond to a declining relative supply of commodities, and this would be represented by a shift to the left of the *RS*

[2] Singer (1950) went further to argue that foreign direct investment had also failed to spread the benefits of technical progress, because it tended to be isolated into enclaves with developing countries, and thus have

schedule to *RS*′. The result would be a shift in the equilibrium from point *A* to point *B* and an increase the relative price of primary commodities. This relative price change would constitute an improvement the terms of trade of commodity exporters (which Prebisch and Singer supposed were developing countries). What we have then is a mechanism, essentially Ricardian in origin, by which technical progress in industrialized countries translates into welfare gains for developing countries.

The main point of Prebish and Singer was that this mechanism didn't work: instead of rising, the relative price of commodities in terms of manufactures had actually fallen. They based this conclusion on a visual inspection of the net barter terms of trade—the relative price of exports to imports—of the United Kingdom from 1876 to 1947. The inverse of this was taken to be a proxy for the relative price of primary commodities to manufactures.

Prebish and Singer also offered theories as why the downward trend had occurred and why it was likely to continue. These can be understood by way of diagram 1 as well. There are essentially two reasons why commodities might experience declining relative prices, despite their lagging technology. One is that somethingelse may prevent the relative supply schedule from shifting to the left or even cause it to shift to the right. The latter would result in an equilibrium at point *D*, with a lower relative commodity price. The second possibility is that something causes the relative demand schedule to shift to the left along with relative supply. If the shift in *RD* is greater than that of *RS*, the result would be an equilibrium like point *C*, again with a lower relative commodity price. Over these two alternative explanations for the decline in commodity prices, one involving supply, the other demand, Prebisch and Singer parted company with each other.

---

few spillovers.

Prebisch offered a supply side theory, based on asymmetries between industrial and developing countries and Keynesian nominal rigidities. The idea was that strong labor unions in industrialized countries caused wages in manufacturing to ratchet upwards with each business cycle, because wages rise during upswings but are sticky during downswings. This, in turn, ratchets up the cost of manufactures. In developing countries, Prebisch argued, weak unions fail to obtain the same wage increases during upswings and cannot prevent wage cuts during downswings. Thus, the cost of primary commodities rises by less than manufactures during upswings and falls by more during downswings, creating a continuous decline in the relative cost of primary commodities, i.e., rightward movement in the relative supply schedule.

Singer focused more on the demand side, considering mainly price and income elasticities. Singer argued that monopoly power in manfuctures prevented the technical progress in that sector from lowering prices, i.e., preventing the leftward shift in RS, much like the argument of Prebisch. However, Singer also argued that the demand for primary commodities showed relatively low income elasticity, so income growth tended to lower the relative demand for, and hence relative price of, primary commodities. Moreover, he argued that technical progress in manufacturing tended to be raw-material saving (e.g., synthetics), thereby causing the demand for primary products to grow slower than for manufactures. Both of these arguments would be reflected in a leftward shift in RD in diagram 1.

Finally, Prebisch and Singer drew policy implications from what they had found. Both argued that as the way out of their dilemma, developing countries should foster industrialization. While they stopped short of advocating protectionism, it is clear that they had in mind to change the pattern of comparative advantage. Thus, whether intentionally or not, Prebisch and Singer provided intellectual support for the import substitution policies that prevailed in many developing countries through the 1970s.

Prebisch and Singer's thesis raises a number of questions that we plan to address in this paper. First, is it reasonable to equate the relative price of commodities with the terms of trade of developing countries in general, and Latin American countries in particular? Second, has the relative price of commodities really declined over the years? Third, are the theories of commodity price determination that Prebisch and Singer put forth plausible? Finally, what policy measures, if any, should developing countries consider toward commodities?

In answering these questions, we shall draw mainly from the literature. However, we shall not attempt a complete review of the literature. For more extensive literature reviews, see Spraos, 1980, Diakosavvas and Scandizzo, 1991, Hadass and Williamson, 2001. Nor will we rely entirely on the literature: in section IV of this paper we offer some new empirical results on the time trend in the commodity terms of trade.

3. HOW IMPORTANT ARE COMMODITY PRICES FOR DEVELOPING COUNTRIES?

Prebisch and Singer assumed that developing countries were specialized in primary commodities and industrialized countries were specialized in manufactures. This generalization led them to treat the relative price of commodities in terms of manufactures as equivalent to the terms of trade of developing countries (and its inverse, terms of trade of industrialized countries). Of course, developing countries do not export only primary commodities, nor do industrialized countries export only manufactures, and thus commodity prices are distinct from the terms of trade. In section, we consider the relevance of this distinction.

The fact that industrialized countries do not export only manufactures was addressed early on by Meier and Baldwin (1957), who pointed out the many primary commodities, like wheat, beef, wool, cotton and sugar, are heavily exported by industrialized countries. Indeed, Diakosavvas and Scandizo note that the developing-country share of agricultural primary commodities was only 30% in 1983, down from 40% in 1955. Yet Spraos (1980) argues that this fact is immaterial, because the same trends that are observed in the broad index of primary commodity prices are found in a narrower index that includes only developing-country products.

How specialized are developing countries in primary commodities? One way to get at this is to measure the share of commodities in developing-country exports. This is not a perfect measure, however, because it will tend to fluctuate along with relative commodity prices. In particular, if commodity prices are declining, then the value share of commodities in a country's exports may fall, even without any changes in that country's export volume. Bearing in mind this limitation, we look at export shares to get sense of the degree of specialization and the products in question.

Table 1 from Cashin, Liang, and McDermott (1999) shows the commodities that account for a large share of the export earnings for various developing countries. The countries that derive 50 percent or more of their export earnings from a single commodity tend to be in the Middle East and in Africa, and the commodity is usually oil. Venezuela is the only such country in Latin America. Several countries receive 20-49 percent of export earnings from a single primary commodity. In Latin America this includes Chile in copper, and several others in bananas and sugar. Still more have primary export revenue shares in the 10-19 percent range.

Table 2 shows the top two exported primary commodities (along with the export shares of these commodities) for several Latin American countries over the last century. Since 1900, the export share of the top two primary commodities has fallen in every country but Venezuela. Even in Venezuela it has fallen since 1950. Today only three countries, Venezuela, Chile, and Cuba, have commodity export shares above 40%. This decline may be simply because of declining commodity prices, but more likely it reflects changing comparative advantage: developing countries are competitive in certain areas of manufacturing, while industrialized countries have moved into the production of services. It may also reflect the effect of import-substitution policies in developing countries over the later half of the century.

Table 1: Commodities with a large share of export earnings in a given country (Based on annual average export shares, 1992-97)			
	50 percent or more of export earnings	**20-49 percent** of export earnings	**10-19 percent** of export earnings
Middle East			
Crude petroleum	Bahrain, Saudi Arabia, Iran, Iraq, Kuwait, Libya, Oman, Qatar, Yemen	Syria, United Arab Emirates	Egypt
Aluminum			Bahrain
Africa			
Crude petroleum	Angola, Gabon, Nigeria, Congo Rep.	Cameroon, Equatorial Guinea	Algeria
Natural gas		Algeria	
Iron Ore		Mauritania	
Copper	Zambia		Congo, Dem. Rep.
Gold		Ghana, South Africa	Mali, Zimbabwe
Timber (African Hardwood)		Equatorial Guinea	Central African Rep., Swaziland, Gabon, Ghana
Cotton		Benin, Chad, Mali, Sudan	Burkina Faso
Tobacco	Malawi	Zimbabwe	
Arabica coffee	Burundi, Ethiopia	Rwanda	
Robusta coffee	Uganda		Cameroon
Cocoa	Sao Tempe and Principe	Cote d'Ivoire, Ghana	Cameroon
Tea			Kenya, Rwanda
Sugar		Mauritius	Swaziland
Western Hemisphere			
Crude petroleum	Venezuela	Ecuador, Trinidad Tobago	Colombia, Mexico
Copper		Chile	Peru
Gold			Guyana
Cotton			Paraguay
Arabica coffee			Colombia, Guatemala, Honduras, Nicaragua, El Salvador
Sugar		Guyana, St. Kitts & Nevis	Belize
Bananas		St. Vincent, Honduras	St. Lucia, Costa Rica, Ecuador
Fishmeal			Peru
Rice			Guyana
Europe, Asia and Pacific			
Crude petroleum		Azerbaijan, Papua New Guinea, Brunei Darussalam, Norway, Russia	Indonesia, Kazakhstan, Vietnam
Natural gas	Turkmenistan		
Aluminum		Tajikistan	
Copper		Mongolia	Kazakhstan, Papua New Guinea
Gold		Papua New Guinea	Uzbekistan
Timber (Asian hardwood)		Lao P. D. R., Solomon Islands	Cambodia, , Papua New Guinea, Indonesia, Myanmar
Timber (softwood)			Latvia, New Zealand
Copra & coconut oil	Kiribati		
Cotton		Pakistan, Uzbekistan	Azerbaijan, Tajikistan, Turkmenistan

Source: Cashin, Liang, and McDermott (1999)

Table 2: Top Two Commodities Exported by Latin American Countries, 1900-1995
(Share of each commodity in total exports, f.o.b.)

	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	1995
Argentina	wool (24) wheat (19)	wheat (23) wool (15)	wheat (24) meat (18)	wheat (19) meat (18)	meat (23) wheat (16)	wheat (17) meat (15)	meat (22) wool (14)	meat (25) wheat (6)	meat (13) wheat (10)	meat (7) wheat (6)	oil (8) wheat (5)
Bolivia	silver (39) tin (27)	tin (54) rubber (16)	tin (68) silver (11)	tin (84) copper (4)	tin (80) silver (6)	tin (67) lead (9)	tin (66) lead (7)	tin (50) gas (16)	tin (43) gas (25)	gas (26) zinc (16)	zinc (11) gas (10)
Brazil	coffee (57) rubber (20)	coffee (51) rubber (31)	coffee (55) cocoa (4)	coffee (68) cotton (3)	coffee (34) cotton (18)	coffee (62) cocoa (7)	coffee (55) cocoa (6)	coffee (32) iron (7)	soya (12) coffee (10)	soya (9) iron (8)	soya (8) iron (6)
Chile	nitrate (65) copper (14)	nitrate (67) copper (7)	nitrate (54) copper (12)	nitrate (43) copper (37)	copper (57) nitrate (19)	copper (52) nitrate (22)	copper (67) nitrate (7)	copper (79) iron (6)	copper (46) iron (4)	copper (46) fish (4)	copper (39) wood (6)
Colombia	coffee (49) gold (17)	coffee (39) gold (16)	coffee (62) gold (13)	coffee (64) oil (13)	coffee (62) oil (13)	coffee (72) oil (13)	coffee (75) oil (13)	coffee (59) oil (13)	coffee (54) oil (13)	oil (23) coffee (21)	coffee (20) oil (19)
Costa Rica	coffee (60) banana (31)	banana (53) coffee (32)	coffee (51) banana (33)	coffee (67) banana (25)	coffee (54) banana (28)	coffee (56) banana (30)	coffee (53) banana (24)	coffee (29) banana (29)	coffee (27) banana (22)	banana (24) coffee (17)	banana (24) coffee (14)
Cuba	sugar (61) tobacco (23)	sugar (70) tobacco (24)	sugar (87) tobacco (10)	sugar (68) tobacco (17)	sugar (70) tobacco (8)	sugar (82) tobacco (5)	sugar (73) tobacco (8)	sugar (75) tobacco (4)	sugar (82) nickel (5)	sugar (74) nickel (7)	sugar (50) nickel (22)
Mexico	silver (44) copper (8)	silver (28) gold (16)	oil (67) silver (17)	silver (15) oil (14)	silver (14) zinc (13)	cotton (17) lead (12)	cotton (23) coffee (9)	cotton (8) coffee (5)	oil (65) coffee (4)	oil (32) coffee (2)	oil (10)
Peru	sugar (25) silver (18)	copper (20) sugar (19)	sugar (35) cotton (26)	oil (33) copper (21)	oil (26) cotton (21)	cotton (34) sugar (15)	cotton (18) copper (17)	fish (27) copper (25)	oil (20) copper (18)	copper (18) fish (13)	copper (19) fish (15)
Uruguay	wool (29) hides (28)	wool (40) hides (23)	wool (40) meat (30)	meat (37) wool (27)	wool (45) meet (22)	wool (48) meat (19)	wool (57) meat (20)	wool (32) meat (32)	wool (17) meat (17)	wool (16) meat (11)	meat (14) wool (9)
Venezuela	coffee (43) cacao (20)	coffee (53) cacao (18)	coffee (42) cacao (18)	oil (82) coffee (10)	oil (88) coffee (3)	oil (94) coffee (1)	oil (88) iron (6)	oil (87) iron (6)	oil (90) iron (2)	oil (79) alumin (4)	oil (75) alumin (4)

Source: Thorp, 1998.

Several studies have taken a more rigorous approach to measuring the importance of commodity prices for the terms of trade of developing countries. Bleaney and Greenaway (1993), for example, estimate a cointegrating regression for non-oil developing countries from 1955-89, in which terms of trade of the developing countries (from IMF data) is expressed as a log-linear function of an index commodity prices and real oil prices. The results show that the series are co-integrated and that for every one percent decline in the relative price of commodities there is a 0.3% decline in the terms of trade of non-oil developing countries. These results are similar to those of Grilli and Yang (1988) and Powell (1991).

By far the most comprehensive study on this topic is Bidarkota and Crucini (2000). They take a disaggregated approach, examining the relationship between the terms of trade of 65 countries and the relative prices of their major commodity exports. Bidarkota and Crucini find that at least 50% of the annual variation in national terms of trade of a typical developing country can be accounted for by variation in the international prices of three or fewer primary commodity exports.

In the final analysis, the importance of commodities in developing countries depends on the precise question one wishes to address. Commodity price trends and fluctuations are clearly important to any policy designed to stabilize commodity prices or the income of commodity producers, such as a stabilization fund or commodity agreement. As noted by Cuddington and Urzua (1989) and Deaton (1992), the effectiveness of a stabilization fund depends crucially on whether shocks to commodity prices are temporary or permanent. Further, an understanding of commodity price trends should also inform longer-term policies affecting the allocation of productive factors across sectors, as was the original intent of Prebisch and Singer.[3] In both of these instances, however, the more disaggregated the data is the better.

Beyond informing policy, Prebisch and Singer sought to use their theory to explain the performance gap between developing and industrialized countries. For this purpose, it is more important to understand the terms of trade of developing countries than to understand commodity prices. This is approach taken by Hadass and Williamson (2001). They bypass the question of the relationship between the terms of trade and commodity prices altogether and simply reexamine evidence on the Prebisch-Singer hypothesis, using country-specific terms of trade data, instead of commodity price data. They construct estimates of the terms of trade for 19 countries, developing and industrialized, and aggregate these into four regions: land-scarce Europe, land-scarce Third World, land-abundant New World and land-abundant Third World. Simply by comparing averages, they find that the terms of trade improved for all regions except the land-scarce Third World. They argue that this is due in part to rapidly declining transport costs during the sample period, which is consistent with Ellsworth's (1956) criticism of Prebisch and Singer.

4. DETERMINANTS OF COMMODITY PRICES: WHAT EXPLAINS THE RELATIVE PRICE OF PRIMARY COMMODITIES?

While most of the literature on the Prebisch-Singer hypothesis has focused on testing the claim of declining relative commodity prices, several papers attempt direct tests of the theories put forth by Prebisch and Singer. Diakosavvas and Scandizzo (1991)

[3] It is not at all clear that a policy of this kind is called for. The point is that, if a policy is to be considered, it

examine Prebisch's theory of asymmetrical nominal rigidities. In particular, they examine the implication that during upswings, the prices of primary products and manufactures should move roughly in tandem, while in downswings, prices of primary products should fall much more than for manufactures. They test this by looking at whether the elasticity of primary product prices with respect to manufactures prices is higher on downswings than on upswings. It turns out that the data reject the hypothesis for all but 5 commodities (nonfood, rice, cotton, rubber and copper).

Bloch and Sapsford (1997, 2000) estimate a structural model to assess the contribution to commodity prices of a number of factors described by Prebisch and Singer. They build a model that assumes marginal cost pricing in the primary sector and mark-up pricing manufactures. Wages are explicitly introduced to try and pick up the effects of unions in manufactures. The model also allows for biased technical change in a la Singer.

Recognizing the potential nonstationarity of the series, Bloch and Sapsford first difference the entire model and apply a two-stage least squares procedure. While this has the intended effect of producing stationarity, it also has the unfortunate effect of sweeping out long-run relationships between the variables. Bloch and Sapsford (1997) find that the main contributing factor to declining commodity prices is raw-material saving technical change. There is also some contribution from faster wage growth in manufactures and a steadily increasing manufacturing markup. The manufacturing mark-up interpretation is suspect, however, as the mark-up is based on price minus labor and intermediate input costs, leaving out rents to other factors, such as capital and land.

Whereas Bloch and Sapsford focus on microeconomic factors affecting commodity prices, Borensztein and Reinhart (1994) and Hua (1998) focus on macroeconomic

---

should be done the information about commodity price trends.

determinants. Borensztein and Reinhart (1994) construct a simple model of the demand for commodities, which depends on the world production of manufactures (commodities are an input into manufacturing) and the real exchange rate of the US dollar (commodities are denominated in dollars, so that an appreciation results in decreased demand for commodities from non-US industrial countries). This is equated with supply, which is treated as exogenous, so that both supply and demand effects can be estimated.

As in Bloch and Sapsford, the model is first-differenced before estimation by GLS. When estimated without the supply component, the model fits well until the mid-1980s after which it over it vastly overpredicts the relative price of commodities. The fit is restored, however, once supply shocks are introduced, and is improved still further after account is taken of the fall in industrial production in Eastern Europe and the former Soviet Union in the late 1980s.

Hua (1998) estimates a demand-side model of commodity prices, similar to that of Borensztein and Reinhart, but he adds in the real interest rate to represent the opportunity cost of holding commodities and lagged oil prices. He estimates the model using a reduced-form error-correction specification. He finds that the hypothesis of a stationary long-run relationship between commodity prices and the levels industrial output and real exchange rate cannot be rejected.

5. EMPIRICAL EVIDENCE ON TRENDS IN PRIMARY COMMODITY PRICES: IS THERE A DOWNWARD TREND IN THE RELATIVE PRICE OF COMMODITIES?

A. Evidence Up through Grilli-Yang (1987)

The bulk of the empirical literature on the Prebisch-Singer hypothesis looks for a secular decline in the relative price of primary commodities in terms of manufactures, rather than directly at the terms of trade of developing countries. Until fairly recently, the largest single obstacle to this search was a lack of good data. Prebisch and Singer had based their conclusions on the net barter terms of the United Kingdom from 1876 to 1947. Subsequent authors criticized the use of these data on several grounds, and various attempts were made to correct for data inadequacies. Spraos (1980) discusses these criticisms in detail (see box for a summary) and also provides estimates based on marginally better data than those used by other authors to that point. Spraos concluded that over the period 1871-1938 a deteriorating trend was still detectable in the data, but its magnitude was smaller than suggested by Prebisch and Singer. When the data was extended to 1970, however, the trend became statistically insignificant. Implicit in this conclusion is the notion that the parameters of the simple time trend model have not remained constant over time. We return to this point below.

Box 1: Bad Data?

Numerous authors criticized Prebisch and Singer's use of British terms of trade data to proxy for relative commodity prices. Here are the four main problems, according to Spraos (1980) (and references therein):

1) Britain's terms of trade were not representative of the terms of trade of industrialized countries on the whole.

2) Industrialized countries export primary commodities also, so the inverse of their terms of trade is bad measure of relative commodity prices.

3) Transport costs: British exports were valued f.o.b. (i.e., without transport costs), while its imports were valued c.i.f. (i.e., inclusive of transport costs). Thus, declining transport costs alone could improve the British terms of trade, thereby overstating the drop in commodity prices.

4) Quality and new products: introducing new manufactured goods and improving the quality of existing ones may push up the price index of manufactures, giving the impression of a decline in the relative price of commodities.

Sapsford (1985) extended the Spraos data, and considered the possibility of a once-and-for-all (or "structural") break in the time trend of relative commodity prices. He showed there to be a significant overall downward trend of 1.3% per year with a large, upward, nearly parallel, shift in the trend line in 1950.

Many of the data issues raised by early authors were put to rest by Grilli and Yang (1988), who carefully constructed a price index of twenty-four internationally traded nonfuel commodities spanning the period 1900-1986. The nominal prices are drawn from a World Bank database consisting annual observations on the twenty-four non-fuel commodities, as well as two energy commodities: oil and coal. The latter are not included in the GY index. The non-fuel group includes eleven food commodities: bananas, beef, cocoa, coffee, lamb, maize, palm oil, rice, sugar, tea, and wheat; seven non-food agricultural commodities: cotton, hides, jute, rubber, timber, tobacco and wool; and six metals: aluminum, copper, lead, silver, tin and zinc. Based on 1977-79 shares, these products account for about 54 percent of the world's nonfuel commodity trade (49 percent of all food products, 83 percent of all nonfood agricultural products, and 45 percent of all metals).

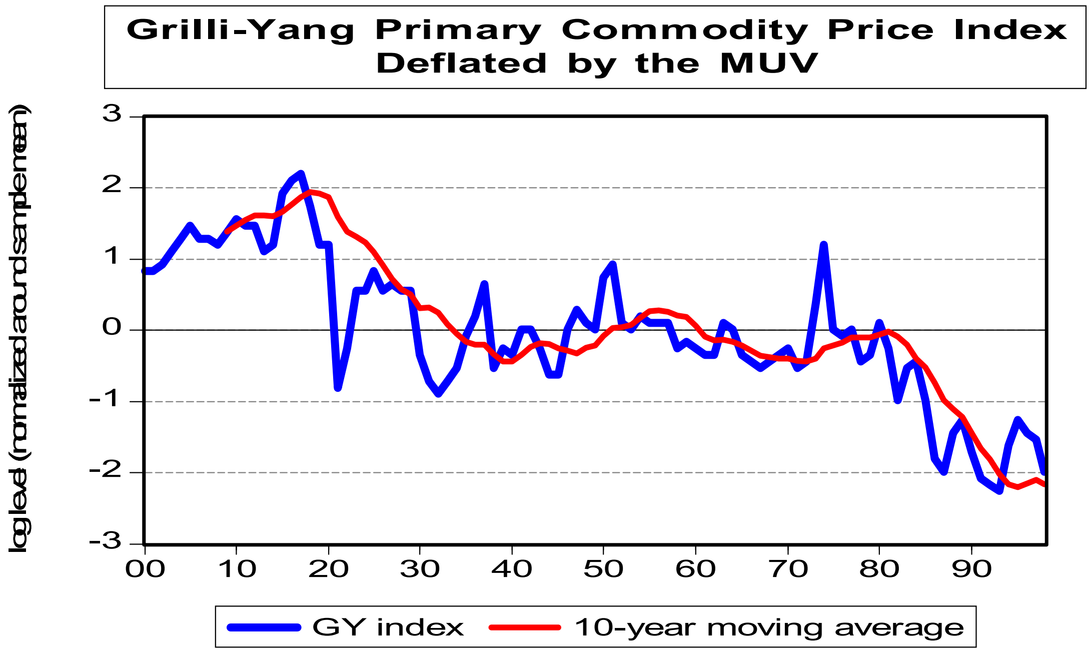
To construct their nominal commodity price index, Grilli and Yang weighted the 24 nominal prices by their respective shares in 1977-79 world commodity trade. To get a real index, GY divided their nominal commodity price index by the a manufacturing unit value index (MUV), which reflects the unit values of manufactured goods exported from industrial countries to developing countries.[4] This is a natural choice of deflators, given PS's concern about the possibility of a secular deterioration in the relative price of primary commodity exports from developing countries in terms of manufacturing goods from the industrial world.

The MUV-deflated GY series, which has recently been extended through 1998 by IMF staff economists, is shown in Fig.1.[5]

[4] GY also considered a U.S. manufacturing price index as a deflator and concluded that their results were not much affected by the choice of deflator.

[5] We thank Paul Cashin of the IMF Research Department for proving the data.

Fig.1



Using their newly constructed index, which covered the 1900-86 period, Grilli and Yang estimated a log-linear time trend and found a significant downward trend of –0.6 percent per year, after allowing for the presence of a downward break in the level the series in 1921. They, therefore, concluded that their findings supported the PS hypothesis.

B. Post Grilli-Yang Work: Econometric Issues

Since the publication of the GY paper and associated long-span dataset in late 1980s, there has been a resurgence in empirical work on long-term trends in commodity prices. The search for a secular trend has shifted from the issue of data quality to econometric issues involved in estimated growth rates or trends in nonstationary time series. Most authors have used the GY dataset, extended to include more recent data in many cases. (In a recent paper, Cashin and McDermott (2001) use The Economist's index of industrial commodity prices covering an even longer time span: 1862-1999 or 140 years! They find a downward trend of -1.3 percent per year.)

Visual inspection of MUV-deflated GY series in Fig. 1, as well as its 10-year moving average, leaves one with the strong impression that it has trended downward over time, as PS conjectured. Modern time series econometrics, however, has taught us that it is potentially misleading to assess long-term trends by inspecting time plots or estimating simple log-linear time trend models. [See the Box: Unit Root Perils.] Although the GY series in Fig. 1 does *not* appear to be mean stationary, it is critical to determine the source of nonstationarity before attempting to make inferences about the presence of any trend. Possible sources of nonstationarity are:

- A deterministic time trend
- A unit root process, with or without drift[6]
- One or more 'structural' breaks in the mean or trend of the univariate process
- General parameter instability in the underlying univariate model.

The key econometric issues are, in short, the possible presence of ***unit roots*** and ***parameter instability*** in the univariate models being estimated. To facilitate a discussion of these issues and to put the existing literature into context, we first specify a general log-linear time trend model that may or may not have a unit root. Second, we describe three types of structural breaks in this framework, where there are sudden shifts in model parameters. A more general type of parameter instability, where parameters are hypothesized to follow random walks, is also considered. Third, we present a chart or matrix showing the relationship among various univariate models that have appeared in the literature. Also included are logical extensions of what has already appeared.

[6] In principle, a series could contain both a deterministic trend and a unit root or more than one unit root; we ignore these cases here.

C. Trend Stationary vs. Difference Stationary Models: Unit Roots

Attempts to estimate the long-term growth rate or trend in an economic time series typically begin with a log-linear time trend model:

$$\ln(y_t) = \alpha + \beta \cdot t + \varepsilon_t \tag{1}$$

In the PS literature, $y=P_C/P_M$ is the ratio of the aggregate commodity price index to the manufacturing goods unit value. The coefficient β on the time index t is the (exponential) growth rate; it indicates the rate of improvement ($\beta > 0$) or deterioration ($\beta < 0$) in the relative commodity price y_t. It is important to allow for possible serial correlation in the error term ε_t in (1). Econometrically, this improves the efficiency of the parameter estimates; economically, it captures the often-pronounced cyclical fluctuations of commodity prices around their long-run trend.

The error process in (1) is assumed to be a general autoregressive, moving average (ARMA) processes:

$$(1 - \rho L) A(L) \varepsilon_t = B(L) u_t \tag{2}$$

It will be convenient in what follows to factor the autoregressive component of the error process in a way that isolates the largest root in the AR part of the error process; this root is denoted denoted ρ. The terms $(1\text{-}\rho L)A(L)$ and $B(L)$ are AR and MA lag polynomials, respectively. The innovations u_t in (2) are assumed to be white noise.

BOX 2: Unit Root Perils

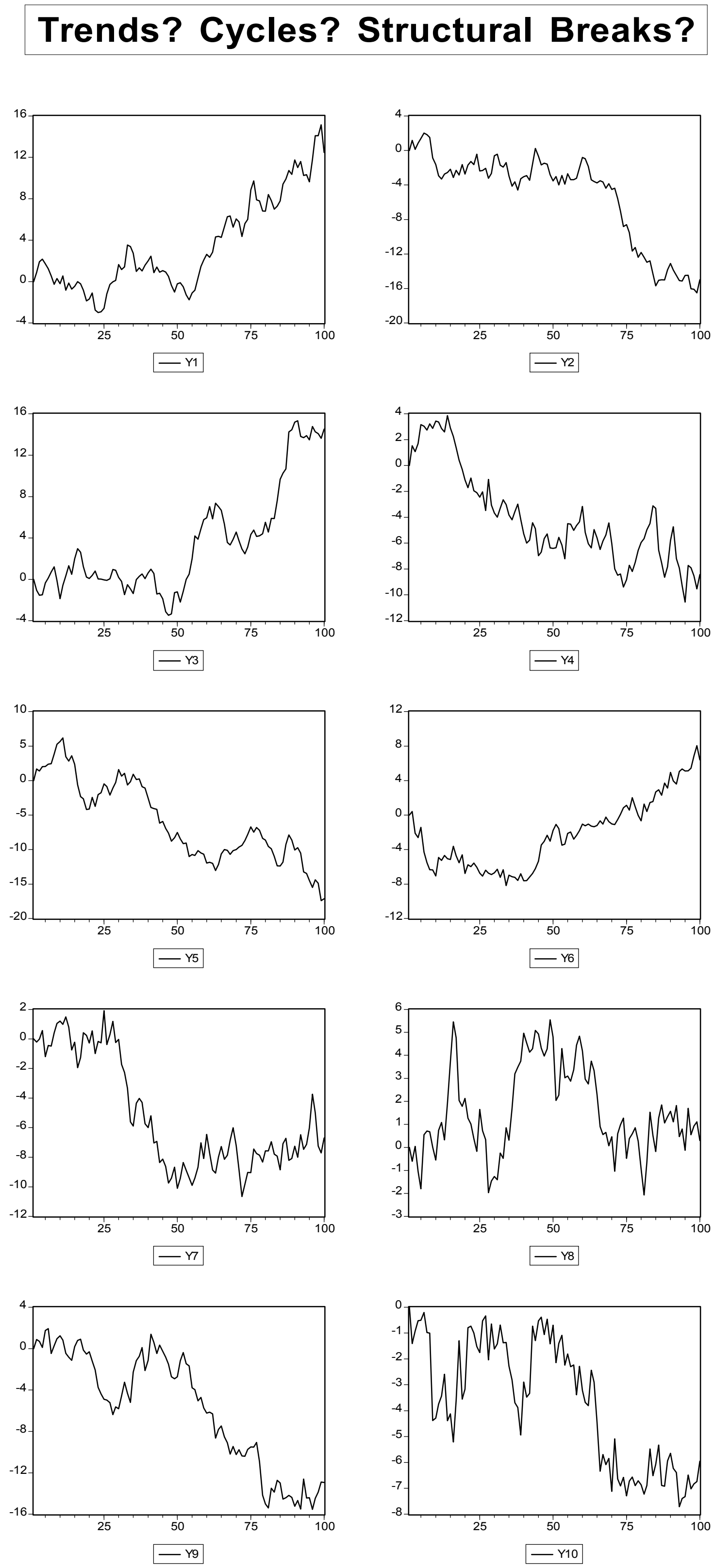
It is now well-known in the time series econometrics literature that attempting to assess long-run trends and detect structural breaks based on graphical evidence and TS models is a highly misleading exercise, especially if the time series are, in fact, unit root process. To illustrate, consider the ten series shown in Fig. 2. Which series exhibit clear positive or negative trends? Which series show structural breaks? Which series have pronounced cyclical behavior?

Reviewing your answers to these three questions, you may find it somewhat surprising to learn that each of the ten series in Fig. 2 is a driftless random walk.[1] So, despite appearances, none of these series has any deterministic trend, cyclical component, or structural break(s)!

Even though these series are really driftless random walks, if you regress each of the series on a constant and a time trend (and correct for apparent first-order serial correlation in the residuals), you will (incorrectly) conclude that nine of the ten series have statistically significant time trends - six are significantly negative; three are significantly positive. This is an example of ***spurious regression*** phenomenon highlighted by Granger and Newbold (1974). There is also spurious cyclicity, reflected in the form of spuriously 'significant' serial correlation coefficients. (See Nelson and Kang (1981).) Moreover, if you eyeball the data to identify dates when there have apparently been structural breaks, then add dummy variables (at the point where visual inspection suggests that the series 'breaks') to your log linear trend models, you will undoubtedly find spuriously significant 'structural breaks' as well.

It is true that visual inspection of the deflated GY series in Fig.1 leaves little doubt that it is nonstationary in the mean, but this need not be the result of a deterministic time trend like (1). The random walk process above is the simplest example of a time series that is nonstationary in the mean due to the presence of a unit root. Unit root processes, with or without, are also nonstationary. The TS and unit root possibilities are nested neatly within the specification in (1)-(2). If $|\rho| < 0$, and $\beta \neq 0$, we have a deterministic time trend model. PS predict $\beta < 0$. If $\rho = 1$ and $\beta \neq 0$, we have a unit root process with drift. Again, if $\beta < 0$, this is consistent with the PS hypothesis. If $\rho = 1$ and $\beta = 0$, we have a driftless unit root process. If real commodity prices are characterized by a unit root, this might be of concern to developing countries or others who specialize or contemplating greater specialization in primary commodities, but not for the reasons PS articulated. The concern should focus on managing risk, rather than coping with secular deterioration.

Fig. 2



A critical issue will be whether $|\rho|<1$, indicating that the error process is stationary, or whether $\rho=1$, indicating nonstationarity due to the presence of a unit over time. In this case, (1)-(2) is referred to a the **trend stationary (TS) model**, indicating that although y_t itself is nonstationary (unless $\beta=0$), fluctuations of y_t around its deterministic trend line are stationary.

If, on the other hand, y_t (or equivalently the error process in (2)) contains a unit root, estimating the TS model – with or without allowance for (supposed) structural breaks – will produce spurious estimates of the trend (as well as spurious cycles). An appropriate strategy for estimating the trend β in this case is to first-difference the model (1)-(2) to achieve stationarity. The result is the so-called **difference stationary (DS) model**, a specification in terms of growth rates rather than log-levels of the y_t series:

$$(1-L)\ln(y_t) \equiv D\ln(y_t) = \beta + v_t \tag{3}$$

where L and D are the lag and difference operators, respectively. The error term in (3) follows an ARMA process:

$$A(L)v_t = B(L)u_t \tag{4}$$

In the DS model, a significant negative estimate of the constant term, β, supports the PS hypothesis.

Using the extended GY dataset (1900-98) to estimate the TS model produces the following estimate:

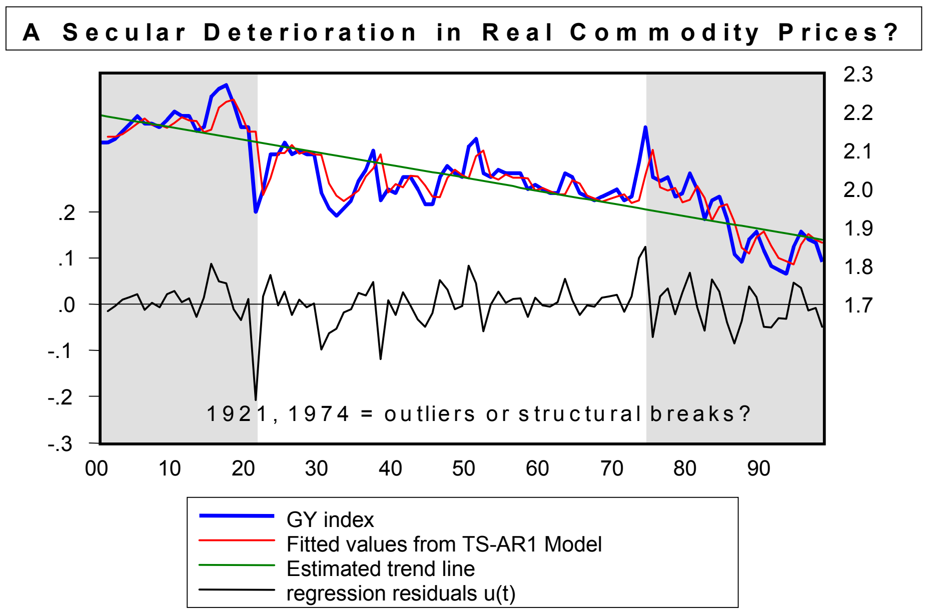
$$y_t = 2.19 - 0.003\cdot t + \varepsilon_t$$

$$\text{where } \varepsilon_t = 0.74\,\varepsilon_{t-1} + u_t$$

The error process is adequately modeled as a first-order AR process. There is a statistically significant trend coefficient equal to –0.3% per year (t=-5.23).

Fitted values from the TS model, the long-term trend estimate, and the regression residuals are shown in Fig. 3. The Figure reveals some potential problems. First, the fitted regression line does not fit the data especially well. Note that the fitted line consistently lags the turning points in the actual data.

Fig. 3



Moreover, the residuals have possible outliers at 1921 and to a lesser extent in 1974 (or 1973). Reexamining the GY series itself, in light of these observations, one might speculate that there have been structural breaks in 1921 and 1974. In subsequent sections of this paper, more formal methods for identifying the timing of a possible break (or two) are considered. These methods indicate clear evidence of a break in 1921, with a second, but statistically insignificant, break in the early 1970s or mid-1980s.

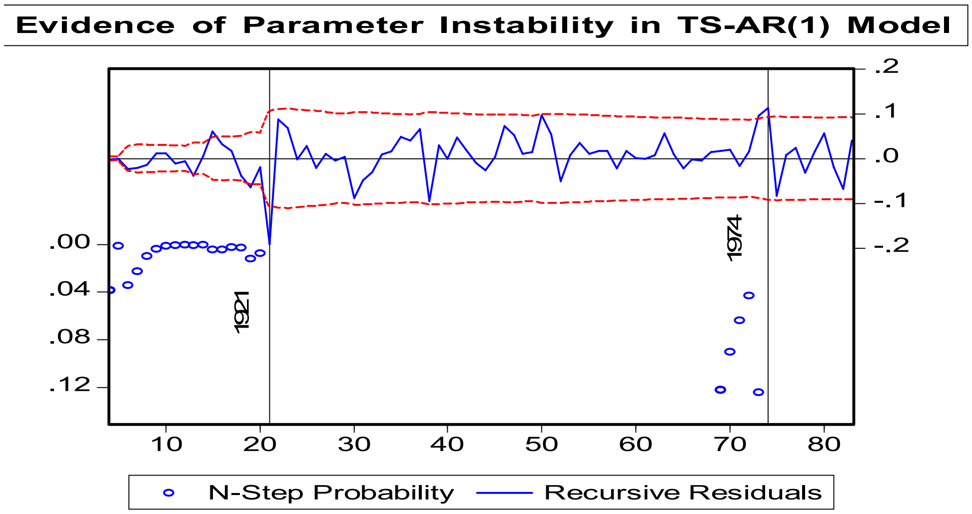
One way to assess the structural stability of the TS-AR(1) model is to calculate recursive residuals and the 2-standard error bands for the hypothesis that the recursive residuals come from the same distribution as the those from the estimated model. As seen in Fig. 4, the recursive residuals in 1921 and 1974 are 'large', suggesting structural breaks.

Figure 4 also shows p-values for an N-step forecast test for each possible forecast sample. To calculate the p-value for 1920, for example, one would use data from 1900 through 1920 to estimate a TS-AR(1) model. This model is then used to forecast y(t) for the remaining N years of the sample: 1921-1998. A test statistic that incorporates the forecast errors, comparing the forecast with the actual value, for the N-steps ahead can be constructed to test the null hypothesis that such forecast errors could have been obtained from the underlying TS-AR(1) model with no structural break. The p-value for the null hypothesis of no structural break gives the probability of finding an even larger test statistic if the null is, in fact, true. If the p-value is smaller than the size of the test, typically .01 or .05, then one should reject the null hypothesis of no structural breaks.

As seen in Fig. 4, the p-values very near 0.00 in the 1910-20 period indicate that the test statistic is so large that the probability of finding a larger one under the null is virtually zero. That is, this graph clearly shows that if the model is fitted with pre-1921 data and used to forecast into the future, there is clear rejection of parameter stability. If instead one uses data up through the 1940s, or 1950s, or 1960s, on the other hand, parameter stability is not rejected. If one uses data through the early 1970s to forecast commodity prices through the end of the 1990s, there is again instability – albeit somewhat less severe judging from the p-values on the left-hand scale in the graph.

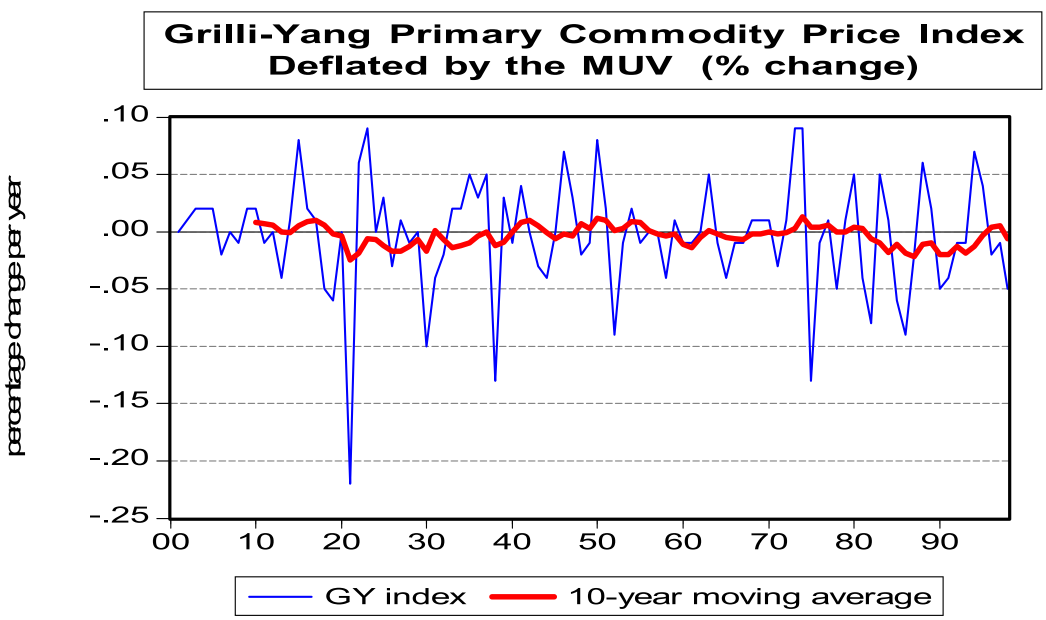
This evidence certainly suggests that the issue of structural breaks or parameter instability must be taken seriously if one chooses the TS model for analyzing the long-term trends in primary commodity prices.

Fig. 4



Consider now the DS model, which uses first-differences of the logged real commodity price series shown in Fig. 5, in order to estimate the growth rate in commodity prices. This specification is appropriate if one believes that the GY series is a unit root process.

Fig. 5: A VOLATILE UNIT ROOT PROCESS?



Note that the $D(y)$ series is very volatile. The 10-year moving average is, not surprisingly, much smoother. It also 'goes through the data' much better than it did the 10-year moving average of the log-levels in Fig.1. This is consistent with the presumption that $D(y)$ is stationary, but y is not. The average value of $D(y)$ is a mere –0.3 percent per year (including the huge –22.0% outlier in 1921). Given the high variance of the series, however, it is not surprising that the null hypothesis of a zero growth rate cannot be rejected.

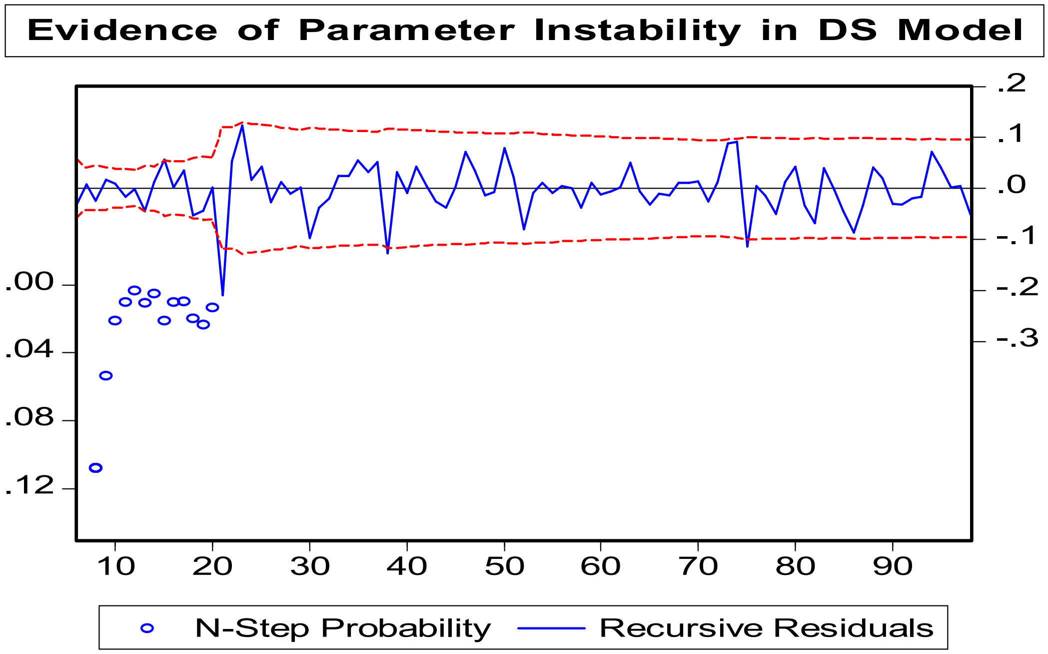
Dependent Variable: DGY
Method: Least Squares
Date: 10/01/01 Time: 17:01
Sample(adjusted): 1903 1998
Included observations: 96 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003996	0.004891	-0.817155	0.4159
DGY(-1)	0.003740	0.100654	0.037160	0.9704
DGY(-2)	-0.258536	0.100665	-2.568285	0.0118

The regression results presented are for a DS model with lagged DGY terms to pick up serial correlation, requiring an AR(2) model to whiten the residuals.

The recursive residual and N-step ahead forecast analysis again suggests that there is a structural break in 1921. See Fig. 6. With the DS model, however, this appears to be the only troublesome episode.

Fig.6



What is clear up to this point? In sum, the possibility of finding statistical significance for the trend in the real GY commodity price index depends critically on whether one believes *a priori*, or concludes on the basis of unit root tests, that GY is trend stationary, or whether it contains a unit root. Regardless of whether the TS or DS specification is chosen, there is evidence of that one or two breaks or parameter instability may be a problem.

<u>D. Structural Breaks and Parameter Instability</u>

It has long been recognized that estimated parameters in models like the TS and DS models above will be biased, or even meaningless, if the true parameters do not remain constant over time. Suppose, for example, that the true growth rate equaled -4.0% in the first half of the sample, but +2.0% in the second half. An econometrician who ignored the shift in parameters might incorrectly conclude that the growth rate was a uniform -2.0 percent over the entire sample.

To consider the possibility of a change in parameters (α, β) in the TS model or β in the DS model,[7] one typically constructs a dummy variable: $DUM_{TB} = 0$ for all $t < TB$ and $DUM_{TB} = 1$ for all $t \geq TB$ where TB is the hypothesized break date. Using this 'level-shift' dummy, as well as its first difference (a 'spike' dummy) and a dummy-time trend interaction term, yields the "TS with break" model and the "DS with break" model, respectively:

TS with Break Model

$$\ln(y_t) = \alpha_1 + \alpha_2 DUM_{TB} + \beta_1 t + \beta_2 (t - TB) * DUM_{TB} + \varepsilon_t \tag{5}$$

DS with Break Model

$$D(\ln(y_t)) = \alpha_2 D(DUM_{TB}) + \beta_1 + \beta_2 * DUM_{TB} + \nu_t \tag{6}$$

These specifications are general enough to encompass the three types of breaks described in Perron (1989) classic paper on testing for unit roots in the presence of structural breaks (which will be discussed below). His model A ("Crash" model) involves only an abrupt shift in the level of the series; i.e. $\alpha_2 \neq 0$, $\beta_2 = 0$. In model B (the 'breaking trend' model), there is a change in the growth rate, but no abrupt level shift: $\alpha_2 = 0$, $\beta_2 \neq 0$. Finally, the 'Combined Model", model C, has change in both the level and growth rate: $\alpha_2 \neq 0$, $\beta_2 \neq 0$.

Suppose that one knows *a priori*, or decides on the basis of unit root testing, whether the TS or DS specification is appropriate. Then, if the break date, TB, is assumed to be known, it is straightforward to test for the presence of structural breaks by examining the t-statistics on α_2 and/or β_2. A test for a break of type C could be carried out using an $\chi 2$ (2) test for the joint hypothesis that $\alpha_2 = 0$ and $\beta_2 = 0$.

---

[7] It is also possible to allow for shifts in the model parameters that describe the error process, its serial correlation and variance, but we do not consider this extension here.

The latter is equivalent to (one variant of) the well-known Chow test for a structural break. More recent work on tests for parameter stability include the above 'structural break' models and associated tests as special cases. See, e.g., Andrews (1993), [8], Ploberger, Kramer, and Kontrus (1989),[9] and Hansen (1992).[10] This literature warns against arguing that the break date TB is known, and hence develops methods for testing for the presence of a possible structural break at an unknown date using an algorithm that searches over all possible break dates.[11]

Recently, there have been attempts in the macroeconomics literature to extend the latter to consider two (possible) break points at unknown dates. (See, e.g. Mehl (2000)). An obvious issue that this extension raises is: why only two breaks rather than, say, three, or four?

---

[8] Andrews (1993) considers tests for parameter instability and structural change with unknown breakpoints in nonlinear parametric models. He tests the null of parameter stability subject to three alternative hypotheses: a one-time structural change either with a known change point, with an unknown change point in a known restricted interval, and with an unknown change point where no information is available regarding the time of the change. The data in the estimated model must be stationary or driftless random walks; they can not be series with deterministic or stochastic time trends. He derives the asymptotic distributions of three test statistics based on the Generalized Method of Moments (GMM) estimators – Wald, Lagrange Multiplier, and Likelihood Ratio-like statistic – under the null hypothesis of constant parameters and provides the respective critical values for each.

[9] Ploberger, Kramer, and Kontrus (1989) propose a 'fluctuations test' for the null hypothesis of parameter constancy over time in a linear regression model with non-stochastic regressors. Their test is based on successive parameter estimates and does not require the location of possible shifts to be known. They derive the asymptotic distribution of the 'fluctuation test' statistic and determine the rejection probability of this test statistic based on the magnitude of fluctuations in the recursive coefficient estimates. They also show how their tests is related to earlier CUSUM and CUSUM squared tests.

[10] Hansen (1992) tests the null of parameter stability in a framework of cointegrated regression models, against the alternative hypotheses that a single structural break exists at either a given or an unknown time. He considers a standard multiple regression model containing I(1) variables that are assumed to be cointegrated; the model parameters are estimated using OLS. His specification also allows for deterministic and stochastic trends in the regressors. He proposes three tests – Sup$\chi 2$, Mean $\chi 2$, and L_C – that test the null hypothesis of parameter constancy and simulates asymptotic critical values for each test. The Sup$\chi 2$ test has greater power against the alternative hypothesis of a one-time break at an unknown date. The mean-$\chi 2$ test has greater power when the alternative is random walk parameters. Interestingly, he shows that the special case of an unstable intercept in under alternative hypothesis can be interpreted as an absence of cointegration among the I(1) variables in the model. Hence his test can be interpreted as a cointegration test where the null hypothesis is the presence of cointegration. (In contract, in the Engel-Granger and Johansen cointegration tests, the null hypothesis is the *absence* of cointegration.)

Authors developing parameter stability tests have also considered the alternative hypothesis where the parameters are assumed to follow a random walk. In this case, the model parameters are generally unstable, in a way that can not be captured a a one-time shift at any particular date. This test of general parameter stability is a good diagnostic test when assessing the adequacy of a particular model specification.

Hansen (1992, p. 321) provides an excellent overview of the issue and possible approaches to dealing with it:

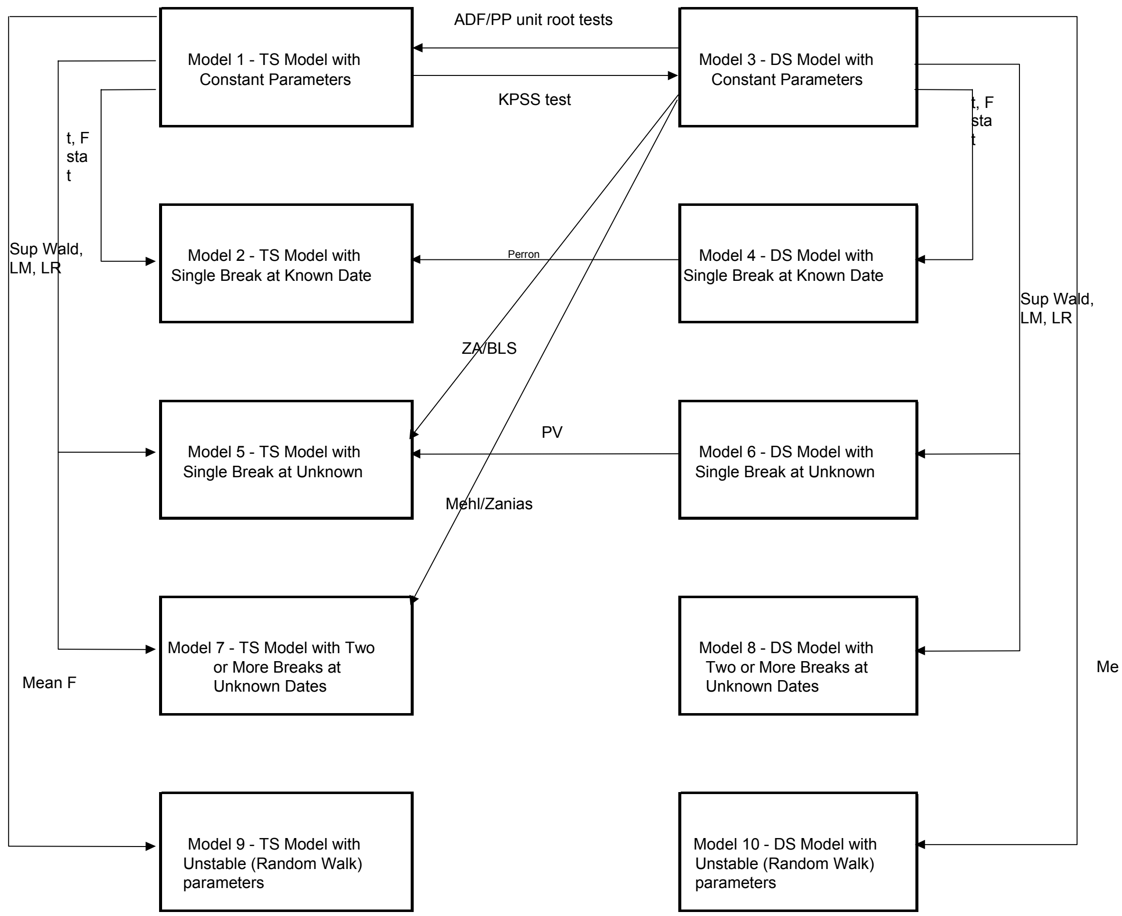
> "One potential problem with time series regression models is that the estimated parameters may change over time. A form of model misspecification, parameter nonconstancy, may have severe consequences on inference if left undetected. In consequence, many applied econometricians routinely apply tests for parameter change. The most common test is the sample split or Chow test (Chow 1960). This test is simple to apply, and the distribution theory is well developed. The test is crippled, however, by the need to specify a priori the timing of the (one-time) structural change that occurs under the alternative. It is hard to see how any non-arbitrary choice can be made independently of the data. In practice, the selection of the breakpoint is chosen either with historical events in mind or after time series plots have been examined. This implies that the breakpoint is selected conditional on the data and therefore conventional critical values are invalid. One can only conclude that inferences may be misleading.
>
> An alternative testing procedure was proposed by Quandt (1960), who suggested specifying the alternative hypothesis as a single structural break of unknown timing. The difficulty with Quandt's test is that the distributional theory was unknown until quite recently. A distributional theory for this test statistic valid for weakly dependant regressors was presented independently by Andrews (1990), Chu (1989), and Hansen (1990). Chu considered as well the case of a simple linear time trend.
>
> Another testing approach has developed in the statistics literature that specifies the coefficients under the alternative hypothesis as random walks. Recent expositions were given by Nabeya and Tanaka (1988), Nyblom (1989), and Hansen (1990).
>
> The preceding works did not consider models with integrated regressors. [Hansen (1992), from which this quote is taken] makes such an extension."

In situations where one is tempted to argue that there are several structural breaks, it probably makes sense to ask whether the situation might be better described a one of general parameter instability.

E. A Matrix of Possible Univariate Specifications and Tests

As outlined above, the key issues in estimating the long-term trend in real commodity prices involve the presence or absence of unit roots and parameter stability. In order to organize our discussion of the existing literature on unit roots and structural breaks, and to point to direction for future research, consider the alternative univariate specifications in chart in Fig. 7. The models in the left column assume that the time series in question, here the real GY commodity price index, does not have a unit root. Rather it is stationary or trend stationary. Those on the right presume the presence of a unit root. Going across the rows, we consider parameter stability/instability of various kinds. The first row assumes the model parameters are constant over time. The second row assumes that there is at most single break or parameter shift in parameters at a known date. The third row assumes the possible single break occurs at an unknown date. The fourth row considers the possibility of two or more breaks – determined by either formal or informal methods where the break dates are known or unknown. Finally, the fifth row considers the case where the model parameters follow a random walk and hence are 'unstable' over time. For convenience the models are numbered for future reference.

Fig. 7: Alternative Specifications



Empirical economists have long employed the TS model for estimating long-term growth rates. A number of these authors also considered the possibility of model 2 – a TS model with a structural break at a known/predetermined date. To formally compare models 1 and 2, Chow-type structural break tests were employed. These tests are represented by the arrow running from model 1 to model 2. The arrow emerges from the model that is assumed to hold under the null hypothesis in the test and points toward the model under the alternative hypothesis.

The unit root revolution in time series econometrics emerged slowly in the mid 1970s and exploded in the 1980s. It stressed that seriously biased (indeed inconsistent) estimates of long-term trends could result if one employed simple log-linear trend models when, in fact, the underlying series had unit roots. Unit root tests, such as those of Dickey and Fuller (1979) and later Phillips-Perron (1988), were proposed as a method for choosing between so-called trend stationary (TS) and difference stationary (DS) models when estimating growth rates or trends in economic time series. The null hypothesis under these tests is the presence of a unit root. These are, therefore, represented by the arrow running from model 3 to model 1. Subsequently, Kwiatkowski, Phillips, Schmidt and Shin [KPSS] (1992) developed a test that maintained mean stationarity or trend stationarity under the null hypothesis. This test is, therefore, represented by the arrow running from model 1 to model 2.

The work of Perron (1989) was seminal in that it demonstrated that the unit root and structural break issues are intertwined. Perron showed how the presence of a structural break at a known break date TB would bias standard unit root tests toward nonrejection of the null hypothesis of a unit root. That is, if one used ADF tests to test model 3 against model 1, when the true model was in fact model 2, one was very likely to falsely accept the null hypothesis of a unit root. This has become known as the 'Perron phenomenon.' Perron went on to develop unit root tests that allowed for the (possible) presence of a structural break under both the null and alternative hypotheses. The Perron-Dickey-Fuller unit root test is represented by the arrow running from model 4 (the null) to model 2 (the alternative). Actually, he developed separate tests for breaks of types A,B, and C, respectively, as described in the accompanying box, Figure 8. [12] The appropriate specification in his various examples was primarily based on eyeballing the data (albeit with some knowledge of post World War I economic history), both to determine the most plausible break date, TB, and the type of break (A,B,C).

Fig.8: Perron's (1992) Model Specification for Carrying Out P-ADF Unit Root Tests in Presence of Break at Time TB[13]

Model A:

$$y_t = \hat{\mu}^A + \hat{\beta}^A t + \hat{\phi}^A DUM_{TB,t} + \hat{d}^A D(DUM_{TB})_t + \hat{\alpha}^A y_{t-1} + \sum_{i=1}^{k} \hat{c}_i \Delta y_{t-i} + \hat{e}_t$$

Model B:

$$y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\phi}^B DUM_{TB,t} + \hat{\gamma}^B DT_t + \hat{\alpha}^B y_{t-1} + \sum_{i=1}^{k} \hat{c}_i \Delta y_{t-i} + \hat{e}_t$$

Model C:

$$y_t = \hat{\mu}^C + \hat{\beta}^C t + \hat{\phi}^C DUM_{TB,t} + \hat{\gamma}^C DT_t + \hat{d}^C D(DUM_{TB})_t + \hat{\alpha}^C y_{t-1} + \sum_{i=1}^{k} \hat{c}_i \Delta y_{t-i} + \hat{e}_t$$

where:

t = time trend and *TB* refers to the time of break.

$DUM_{TB,t} = 1$ if $t \geq TB$, and 0 otherwise	(level-shift dummy)
$D(DUM_{TB})_t = 1$ if $t = T_B$, and 0 otherwise	(spike dummy)
$DT_t = (t-TB)*DUM_{TB,t}$	(time-interaction dummy)

[12] Perron's work on structural breaks distinguishes between the *Additive Outlier Model* and the *Innovational Outlier Model*. In the former, the break occurs suddenly at the break date. In the latter, the break takes the form of a shift in the structure of the underlying model that takes effect gradually over time in exactly the same way that an innovation is perpetuated by the ARMA process of the estimated model. See Perron and Vogelsang (1992) for a discussion of the two models. Throughlut this paper, we use the innovational outlier model.

[13] This is slight reworking of Perron's original specification in that the timing of the dummy here reflects the first period of the new regime and the time interaction term is written the same way in models B and C. This shows more clearly that models A and B are nested in C.

Table 3 shows how imposing restrictions on the test equation for model C above causes it to collapse to TS or DS models with various break types. Unrestricted, the model nests all of these as special cases.

Table 3

Model	α**(ADF stat)**	***d*D*(*DUM*)**	ϕ****DUM***	γ****DT***
DS-no break	0	0	0	0
TS-no break	$\neq 0$	0	0	0
DS-break A	0	$\neq 0$	0	0
TS-break A	$\neq 0$	0	$\neq 0$	0
TS-with single outlier	$\neq 0$	$\neq 0$	0	0
DS-break B	0	0	$\neq 0$	0
TS-break B	$\neq 0$	0	0	
DS-break C	0	$\neq 0$	$\neq 0$	0
TS-break C	$\neq 0$	0	$\neq 0$	$\neq 0$

Subsequent authors, notably Christiano (1992), Banerjee-Lumsdaine-Stock (BLS) (1992) and Zivot and Andrews (ZA) (1992) were highly critical of Perron's assumption that the date of the (possible) break was either known *a priori* or was determined by inspecting the data without adjusting the critical values in subsequent statistical tests to reflect this informal 'search' procedure. This, of course, echoed concerns in the literature developing formal tests for parameter stability (discussed above; see Hansen (1992) quotation). As Fig. 2 illustrates, unit root process often exhibit *apparent* breaks even when, in truth, there is none. So it's risky to assess the presence of breaks by eyeballing the data.

BLS and ZA proposed a generalization of Perron-Dickey-Fuller (P-ADF) test that treated the possible break date as unknown; they propose an algorithm for searching over all possible break dates within the (trimmed[14]) sample. There are a couple of noteworthy aspects of this test, which we dub the ZAP-ADF test. First, it allows the structural break under the alternative hypothesis but not under the null hypothesis of a unit root. This is reflected in the arrow representing the ZAP-ADF test, which runs from model 3 to model 5 in Fig. 7. Second, the ZAP-ADF test is a test of the null hypothesis of a unit root, conditional on the possible presence of structural break at an unknown date. It is not a test for the presence of structural break (hence our phrase 'a *possible* structural break'). In spite of this, the ZAP-ADF and P-ADF tests have repeatedly been represented as tests of structural change in both the applied macroeconometric and commodity price literatures. [See, e.g., Enders (1995), Leon and Soto (1997), and Zanias (undated).] Finally, the ZAP-ADF test assumes that the *type* of break is known *a priori*.[15] Thus, the ZAP-ADF test has the rather inconsistent feature of testing for a unit root, conditional on the possible presence of a *known* type of structural break (A,B,C) at an *unknown* date!

In contrast to the ZAP-ADF test, the specification in Perron (1989) permitted the break under *both* the null and alternative hypotheses -- albeit at a known date. Perron and Vogelsang (1992) developed a unit root test that allowed for a break at an unknown date under both the null and alternative. However, this was done in the context of comparing a TS model with zero trend (β=0) to a DS model (with β=0 here, as well). This, in effect,

[14] For technical reasons, it is often necessary to 'trim' the first and last 10-15% of the sample, so that only break dates in the middle 70-80% of the sample are considered.

[15] That is, while ZA criticize Perron (1989) for assuming the timing of the break is known, they accept his visual characterization of the most plausible type of break for each macroeconomic variable considered as they demonstrate how their test differs from his.

limited the analysis *ex ante* to breaks of type A. This test is denoted PV, running from model 6 to model 5 in the matrix.

More recently, Leybourne, Mills, and Newbold (LMN)(1998) have pointed to a very compelling reason for preferring unit root testing procedures that allow for the presence of a break under both the null and alternative hypotheses. They consider situations where the true model is a DS model with either a type A[16] or B break. In either case, (LMN (1998, p.191), our emphasis) demonstrate that there is a 'converse Perron phenomenon.'[17] Specifically, "if the break occurs early in the series, routine application of standard Dickey-Fuller tests can lead to a very serious problem of spurious *rejection* of the unit root null hypothesis." They go on to emphasize that:

> Of course, this problem will not occur when the test procedures that explicitly permit a break under the null as well as under the alternative are employed, as for example in Perron (1989, 1993, 1994) and Perron and Vogelsang (1992). This is the case whether the break date is treated as exogenous or as endogenous, as in Zivot and Andrews (1992) or Banerjee et al. (1992). Indeed, our results imply a further motivation for employing such tests when a break is suspected, in addition to the well-known lack of power of standard Dickey-Fuller tests in these circumstances. (1998, p.198)

As mentioned above, some authors have entertained the possibility that economic time series might have more than one structural break. For the most part these multiple breaks were identified by casual data inspection, although there are now formal unit root tests in the (possible) presence of two structural breaks at unknown dates. See, e.g., Mehl (2000). Unfortunately, the unit root tests in the latter paper shares two undesirable features

---

[16] They note that in the case of their type A break in the simplest type of DS model – a driftless random walk, this implies that the first-difference of the series is white noise with a single outlier at the break date.
[17] LMN (1998, p.191): "It is well known that if a series is generated by a process that is stationary around a broken trend, conventional Dickey-Fuller tests can have very low power. [i.e., the 'Perron phenomenon.'] In this paper, the converse *phenomenon* is studied and illustrated. Suppose that the true generating process is integrated of order one, but with a break…"

of earlier work: the breaks are assumed to be of a known type and the break is allowed under the alternative hypothesis, but not under the null hypothesis of a unit root.

F. A Selective Review of Post Grilli-Yang Empirical Work

As mentioned earlier, the literature through Grilli-Yang (1988) used the TS model – model 1 in Fig.3 -- to estimate the long-term trend in real commodity prices. A number of these authors recognized the possibility of structural changes in the form of one-time shifts in the level and/or trend in the real commodity price series. That is, they compared models 1 and 2. For example, Sapsford (1985) found a break in 1950 using pre-GY data, as mentioned earlier. GY(1987) and Cuddington- Urzúa (1989) both identified a breakpoint in 1921 using the GY dataset for the period 1900-1983. Contrary to GY, CU showed that, after accounting for the highly significant downward shift in the level of the real GY price index in 1921, the trends on either side of the break were not significantly different from zero in the TS specification. Not surprisingly, if one ignored the one-time downward step in the data, the estimated trend coefficient β was negative and significant. This illustrates the potential for incorrect statistical inferences if structural shifts are ignored.

CU also demonstrated that the structural break in 1950 detected by Sapsford (1985) using the trend stationary model on pre-Grilli-Yang data, was not significant when using the GY data once the 1921 break was included.

Cuddington and Urzua (1989) were the first to carry out unit root tests on the GY commodity price index. They were unable to reject the unit root hypothesis, and therefore, stated a preference for DS models rather than TS models when estimating the long-term trend in real commodity prices. Using data from 1900-83, they were unable to reject the null hypothesis that $\beta=0$ in the DS model in (3)-(4). This was the case whether or not they allowed for a one-time drop in the level of the GY series in 1921. In terms of Fig. 7, CU (1989) considered models (3) and (4), and formally tested model 3 against model 1 (assuming no unit root) and model 3 against model 4 (assuming there is a unit root). By carrying out ADF tests, they compared model 3 to model 1, and using a new unit root test in Perron (1989), they tested model 4 against model 2.

Applying now-standard augmented Dickey-Fuller unit root tests as well as Perron-ADF tests that allow for a possible structural break at a predetermined break date, CU (1989) showed that the unit root hypothesis can not be rejected for the GY index. When CU estimated the DS model using GY data from 1900-1983, the estimated long -term growth rate was statistically insignificant, regardless of whether or not one included a spike dummy to account for the downward shift in the level of the real GY series in 1921.

The DS specification, therefore, leads to the conclusion that real commodity prices follow a driftless unit root process. The policy implications from this specification are quite different from those based on the CU's TS model with a one-time level shift in 1921. The risk entailed for commodity producers, exporters, and commodity stabilization fund managers is considerably greater if one believes that the true model is the DS specification. The CU unit root tests failed to reject the null hypothesis of a unit root, but such tests have notoriously low power so no definitive conclusion is warranted.

Note that the DS model with a one-time level shift in 1921 is a very plausible candidate model for the GY series. In fact, it is the specification preferred by Cuddington-Urzua (1989). The year 1921, moreover, occurs early in the sample, precisely the situation where LMN warn that DF tests (or ZAP-ADF tests that do not allow for a break under the null) are likely to lead to false rejections! In spite of this bias, CU did not reject the unit root when they assumed a known break date. Assuming an unknown break date implies smaller (i.e. more negative) critical values for the resulting ZAP-ADF test. So, again, one would expect not expect to reject the unit root hypothesis.

Cuddington (1992 JDE) repeated the exercise of testing for unit roots (with or without breaks at possible break dates determined by visual inspection) for each of the 24 component commodities in the GY index (1900-1983). Some commodities had unit roots; others did not. Some commodities had negative price trends, while others had positive trends. Surprisingly, not a single commodity had a structural break in 1921! [18] This led Cuddington-Wei (1992) to conjecture that there was some time aggregation issue involved in the construction of the GY index, as theirs was an arithmetic index. Cuddington-Wei construct a geometric index, so that the results from the individual commodities should be reflected in the geometric index, as it is just a simple weighted average of the logs of the individual commodity prices that comprise the index. Using the Cudd-Wei index (over the slightly extended period 1900-1988), they find that unit root tests are inconclusive. The estimated trend in the real commodity price index, however, turns out to be statistically insignificant regardless of whether one uses the TS or DS model specification.

Subsequent work has reconsidered Cuddington and Urzúa's claim of a trendless series with a break in 1921. Powell (1991) for example found three downward jumps, in 1921, 1938 and 1975, and no continuous trend. Ardeni and Wright (1992) used a "trend plus cycle model" and extend the Grilli-Yang data to 1988 to find a continuous trend of between –0.14% to –1.06%, depending on the exact model specification. Moreover, this trend survives with or without a structural break in 1921. Bleany and Greenaway (1993) avoid the issue of a structural break in 1921, by considering 1925-91 data, and instead find a downward jump in 1980, with no continuous trend.

Leon and Soto (1997) and Zanias (undated) apply the ZA/BLS method for testing for unit roots in the presence of a single break at an unknown break point. Zanias, in particular, finds that this method identifies 1984 as the primary break point. It is, however, difficult to know how to interpret a break point in a portion of the sample that Andrews and others recommend should be trimmed off, because it is too close to the end of the sample. Zanias goes on to re-apply the ZA/BLS approach to find a second break, conditional on the presence of the first break in 1984. This sequential procedure chooses 1921 as the second break point

Although the PS literature has extensively explored the possibility of structural breaks, the more general phenomenon of parameter instability has, to date, been overlooked. [19] This paper makes an initial effort at this extension. Apart from the econometric issues raised by, e.g., Hansen quote above, parameter instability has interesting implications for testing the PS hypothesis. PS did not claim that the long-run trend would necessarily remain constant over time, only that it would be negative!

G. A New Look at Growth Rates, Possible Breaks and Unit Root Tests

In testing the PS hypothesis, our primary interest is in the growth rate β in the deflated GY index. Has it been negative as PS predicted? Has it been relatively stable for time? Or

---

[18] Cuddington found breaks for only coffee (1950) and oil (1974), which is not in the GY index.

[20] When searching for two break points with the use of spike, level-shift and trend interaction dummies, it is easy to show that the break points must be separated by a minimum of two periods to avoid perfect multicollinearity among the dummies. If one does not allow for breaks under the null hypothesis, only under the alternative (as in Mehl (2000)), then the two spike dummies are omitted and the two break need only be separated by a single period to avoid perfect multicollinearity.

has this parameter shifted or drifted over time, or exhibited a sharp structural break or breaks? In our particular application, we are less interested in the presence or absence of unit roots *per se* than was the applied macroeconometric literature. Unfortunately it is difficult to estimate the growth rate β without first making a decision on the presence or absence of a unit root first. Ideally, we would also like to formally test for the presence of structural breaks without prejudging the case of whether the series has a unit root. This objective, however, appears to be beyond our reach at this time.

Our strategy is to proceed as follows. First estimate augmented ZAP-ADF-like regressions allowing for at most two structural breaks at unknown dates. Having searched for the two most plausible break dates, we then test whether each break is statistically significant. If both breaks are significant, we assume two breaks in what follows. If only one break is statistically significant, we re-estimate the ZAP-ADF equation with a single break at an unknown date and test the to see whether the remaining break is statistically significant.

i. The Possibility of At Most Two Break Points

We first consider the possibility that the GY series is characterized by (up to) two structural breaks of unknown type (A,B,C) and at unknown dates. Our search algorithm considers all possible pairs of break dates (TB1, TB2) in the trimmed sample.[20] For the ZAP-ADF equation, three dummies – the spike, level-shift, and trend interaction dummies -- are included for each of the two hypothesized break dates in order to allow for breaks of type A,B, or C under both the null hypothesis of a unit root and the alternative hypothesis of trend stationarity. That is, the estimated ZAP-ADF equation is:

$$y_t = \hat{\mu} + \hat{\beta} t + \hat{\alpha} y_{t-1} + \hat{d}_1 D(DUM_{TB1})_t + \hat{\phi}_1 DUM_{TB1,t} + \gamma_1 * t * DUM_{TB1,t} + \\ + \hat{d}_2 D(DUM_{TB12})_t + \hat{\phi}_2 DUM_{TB2,t} + \gamma_2 * t * DUM_{TB2,t} \sum_{i=1}^{k} \hat{c}_i \Delta y_{t-i} + \hat{e}_t \tag{7}$$

In each regression as different pairs of break dates (TB1, TB2) are considered, the number of lags of the dependent variable, k, is chosen using Perron's general to specific method so as to be reasonably confident that the residuals are serially uncorrelated at each stage as we proceed.

Extending Hansen (1992), albeit less rigorously at this point, to cover situations with two break dates, we calculate an supχ^2 statistic to make an inference about the existence of structural change and a meanχ^2 statistic to determine the existence of general parameter instability in the data. In this context of the ZAP-ADF equation, the supχ^2 statistic is the maximum value over all (TB1, TB2) pairs of the Wald test statistic for the null hypothesis that all six dummies (level, spike, and time interaction dummies for TB1 and TB2) are equal to zero. Hence we will call it a supχ^2(6) statistic. The meanχ2(6) statistic is simply the average of the χ2(6) statistics. As explained in Hansen (1992), a significantly high supχ2 with a relatively low meanχ2 implies the existence of a single structural break (or here two structural breaks) and no/low parameter instability. On the other hand, a high meanχ2 is indicative of general parameter instability rather than an abrupt structural change (or two). In addition, we also compute the χ2 statistic for to test the joint significance of the three types of dummies associated with each candidate break point. These are denoted *χ2 stat(3)_TB1* and *χ2 stat(3)_TB2*, respectively. See Table 4 for results.

Table 4. Grid Search Results for Two Structural Breaks[21]

Type of Model Type of Structural Break Dummies	**ZAP-ADF Model** With Level, Spike & Time Interaction Dummies
Chosen Break Points TB1 & TB2	**1921 & 1974**
Supχ2(6)	58.53
Meanχ2(6)	10.79
ADF stat for unit root test at (TB1=1921,TB2=1974)	-5.93
χ2stat(3)_TB1 (1921)	14.76
χ2stat(3)_TB2 (1974)	7.77

According to the grid search based on the ZAP-ADF equation, the two structural breaks are most likely to have occurred in 1921 and 1974.[22] The sup $\chi^2(6)$ statistic of 58.53 is presumably statistically significant (given that the 1% critical value from the $\chi^2(6)$ distribution is 16.81. The critical value for the sup statistic must be determined via simulation methods, but we know it will be higher than 16.81.) The mean $\chi^2(6)$ value of 10.79, on the other hand, is probably not statistically significant. (We know that for models with one break the critical values from the mean $\chi^2(6)$ distribution will be slightly *lower* than those from the standard $\chi^2(6)$ distribution. See Hansen , 1992.)

Fig. 9 shows a 3-D graph of the χ2 (6) values corresponding to alternative break date pairs. The sup$\chi^2(6)$ of 58.53 corresponding to (TB1,TB2)=(1921, 1974) is, by definition, the global maximum but there are several local maximums. There are, in fact, others χ2(6) statistics that are close in value to the supχ2 attained in (1921, 1974). The second highest supχ2 value of 56.33 occurs with candidate break date pair (1921, 1973),

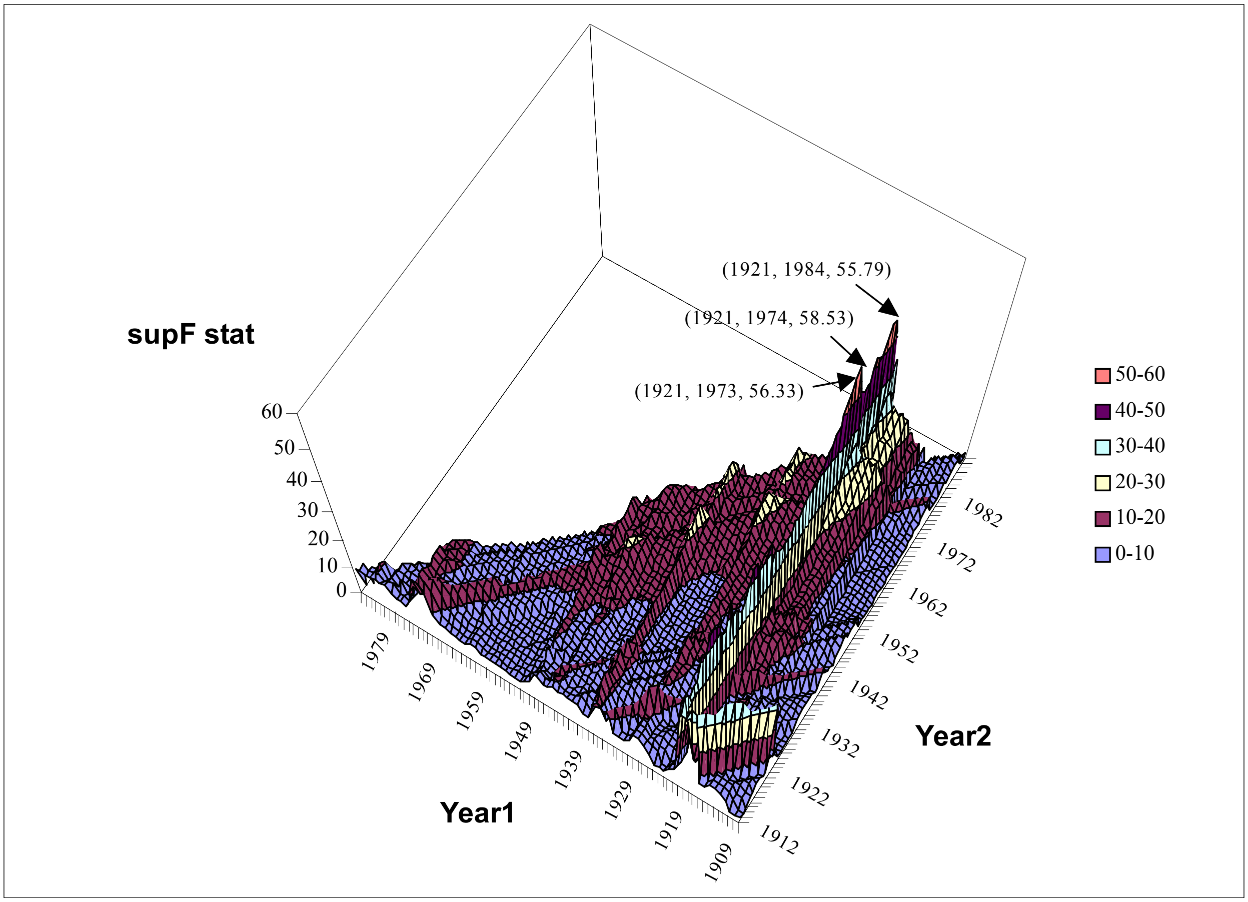
[21] In a Pentium III processor, the program runs for approximately 30 minutes for the ZAP-ADF model. The maximum number of lags considered in the lagged dependent variable polynominal is six.

[22] We also searched for two breaks in the ZAP-ADF model without the two spike dummies (which precludes a level-shift break under the null hypothesis of a unit root). In this specification the most prominent breaks are in 1921 and *1985*. This estimation produced a supχ^2 of 34.43, a meanχ^2 of 8.07, and an ADF stat of –7.29.

and the third highest of 55.79 occurred at (1921, 1984). Note that there is a clear L-shaped 'ridge' of high sup$\chi2(6)$ values where either TB1 or TB2 is 1921. This suggests that there is a rather decisive break in 1921. Placing the other possible break date almost any other date after 1923 in the trimmed sample often produces a high $\chi^2(6)$ statistic. This might be indicative of general parameter instability, rather than a second decisive break point. Alternatively, there may be only a single break at 1921, with the dating of a second possible break being rather inconsequential in determining the value of the sup $\chi^2(6)$ statistic.

Turning to the two break points, considered separately, the $\chi^2(3)$_TB1 and $\chi^2(3)$_TB2 suggests that the structural change in 1921 is more prominent than the one in 1974. Note that $\chi^2(3)$_TB2=7.77, which is less than the standard 1% critical value for $\chi^2(3)$ of 11.34. The appropriate critical value, given that the break dates are chosen from search process that maximizes $\chi^2(6)$, must be higher. Thus, we can safely conclude that TB2 is insignificant. A determination on TB1 would require a calculation of the appropriate critical values. A complementary approach is to estimate the ZAP-ADF equation with a single possible break point at an unknown date, which we take up next.

Fig. 9. 3-D Graph of the $\chi 2(6)$ Statistic for the ZAP-ADF Equation



To reiterate, without conducting an extensive Monte Carlo simulation analysis we don't know whether the sup$\chi 2$ or mean$\chi 2$ statistics are statistically significant. Similarly, we don't know whether the ZAP-ADF stat of –5.93 in Table 4 above is large enough to reject the null hypothesis of a unit root, conditional on the possible presence of two breaks of unknown type (A,B,C) and unknown dates.

ii. The ZAP-ADF Tests with At Most One Break

The above exercise is repeated assuming, now, that there is at most one break at an unknown date as in ZA/BLS and Perron-Vogelsang.[23] The ZAP-ADF equation is:

$$y_t = \hat{\mu} + \hat{\beta}t + \hat{\alpha}y_{t-1} + \hat{d}D(DUM_{TB1})_t + \phi DUM_{TB1,t} + \gamma t * DUM_{TB1,t} + \sum_{i=1}^{k} \hat{c}_i \Delta y_{t-i} + \hat{e}_t \quad (8)$$

ZA/BLS and Perron-Vogelsang (1992) choose the break date that minimizes the t-statistic on α^c – the ADF statistic. We the use an alternative search algorithm, although using our output it is easy to compare to the ZA/BLS results. The alternative we consider is similar to Andrews (1992) and Hansen (1992). For each and every possible break date TB in the [.15, .85]-trimmed sample, we calculate the Wald χ^2 statistic for the joint hypothesis that the coefficients on all three break dummies are jointly insignificant. That is, H_0: $\hat{\theta} = \hat{\gamma} = \hat{d} = 0$.[24] Under the null, there is no break of Type A, B, or C.

We plot the sequence of χ^2 statistics, as in Hansen (1992), to get some indication of whether there might be one or more breaks. The maximum in the sequence of χ^2 statistics, denoted "sup χ^2" is determined. The mean χ^2 is also calculated.[25] A high value for the sup χ^2 statistic signals a possible structural break (of type A, B, or C); a high value for the mean χ^2 statistic, on the other hand, suggests the parameter estimates (in the ZAP-ADF equation

[23] Perron and Vogelsang also consider an algorithm that selects TB so as to maximize the absolute value of the t-statsitic on DUM. In their context which precludes breaks in the growth rate (as it is identically zero), this amounts to using the sup$\chi 2$ statistic that we employ.

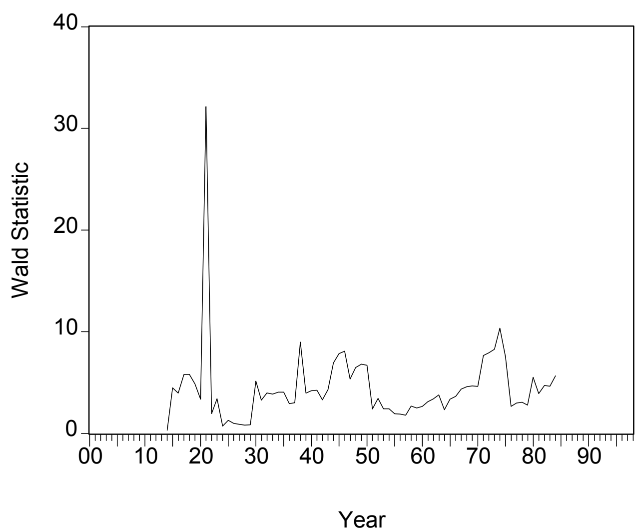
[24] Given that we impose linear restrictions, the Wald test output produces both an $\chi 2$ -statistic and a Chi-square statistic. However, the Chi-square statistic is more appropriate since lagged dependant variables appear as regressors in our equation specification.

[25] These two statistics are analogous to the sup$\chi 2$ and mean$\chi 2$ statistics discussed in Hansen (1992), for his regressions that did not involve lagged dependent variables.

in this case) are unstable. We also plot the t-statistics on the dummies and the ADF-t statistic for each possible break point.

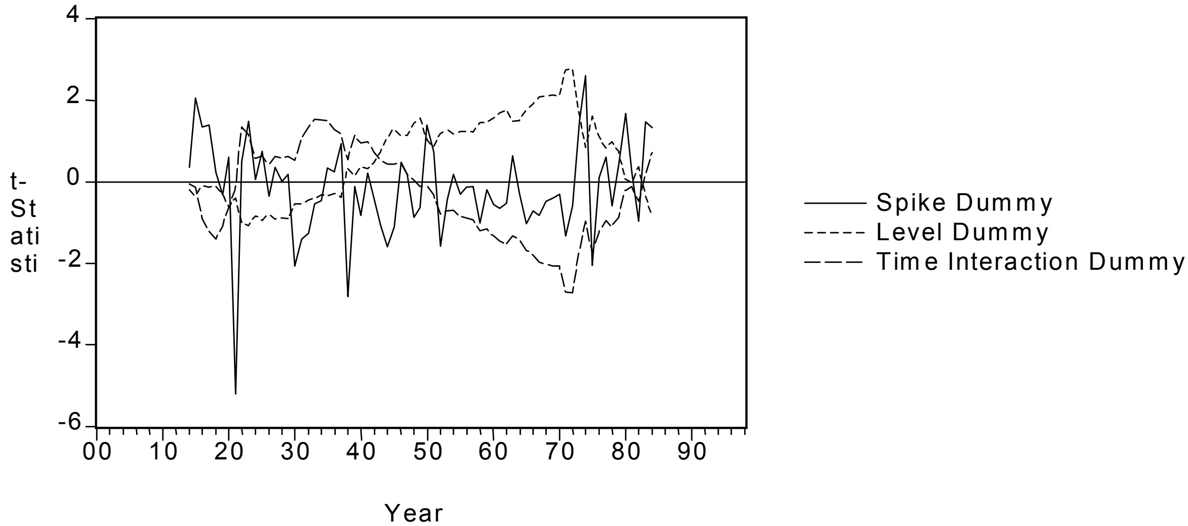
When our single break selection procedure is applied to the deflated GY index, the Wald test statistics for the various possible break points are those shown in Fig. 10. The sup χ^2 of 32.14 occurs in 1921 and is a clear outlier in terms of magnitude; the mean χ^2 = 4.19. Given that sup χ^2 lies well above the standard 1% critical value for χ^2 (4) of 13.28, and mean χ^2 lies well below the critical value, it is reasonable to conclude that the real GY series is well characterized by a single break in 1921, rather than multiple breaks or general parameter instability.

Fig. 10: The Sequence of Wald χ^2 Test Statistics for the Joint Hypothesis H_0: $\hat{\theta} = \hat{\gamma} = \hat{d} = 0$.



To get a better understanding of what is producing the large sup χ^2 value in 1921, one can examine the sequence of t-statistics on the individual dummy coefficients shown in Fig. 11 below.

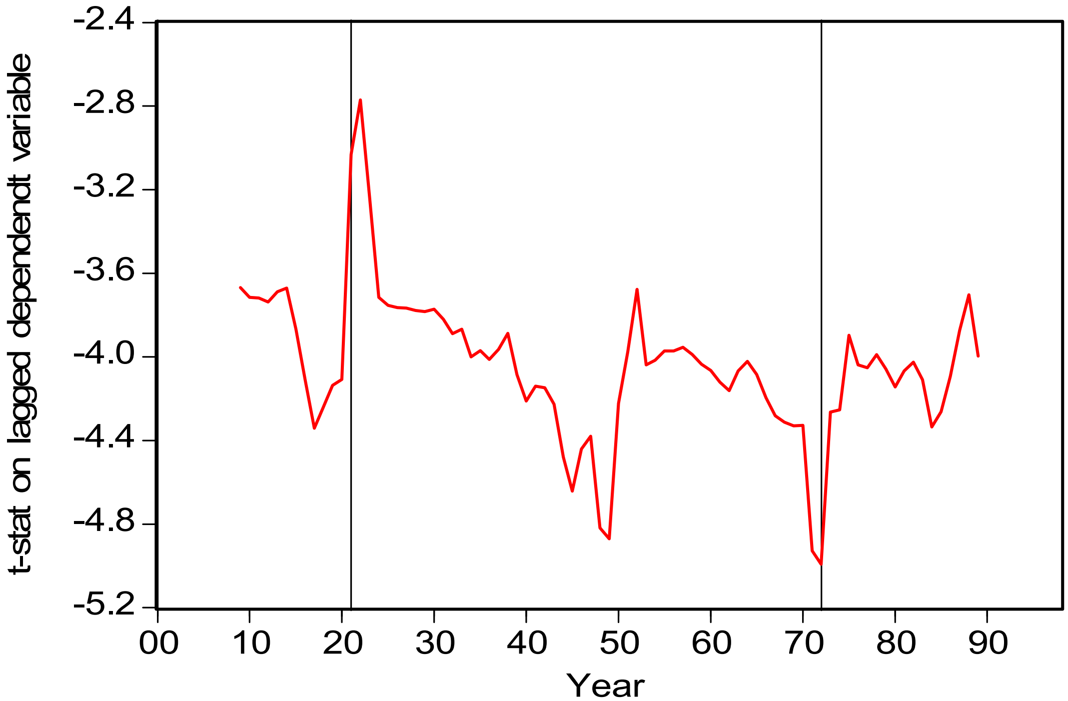
Fig.11



A visual inspection confirms the existence of a spike dummy in 1921. Formally, the t-statistics in 1921 are –5.2047, -0.3987, -0.2141, and –0.2029 for the spike, level, interaction, and trend dummies respectively. Even though we do not have the correct critical values to interpret the spike dummy at this point, a t-statistic of –5.2047 is presumably above the appropriately calculated critical value, implying a rejection of the hypothesis of a zero coefficient on the spike dummy.[26]

[26] The coefficient on the spike dummy in 1921 is –0.2184. This turns out to be a clear outlier compared with that of the rest of the period.

Fig. 12: The Perron-ADF t-statistics at various break dates



Turn now to Fig.12, which shows the ADF t-statistic for all possible (single) break dates. Our supχ^2 statistic identified 1921 as the year of the break. On that date, the ADF statistic has a value of –3.03. Presumably (awaiting correct critical values), the null of unit root cannot be rejected at a reasonable level of significance. Given that Fig.11 shows only the t-statistic on the spike dummy is large, and the ADF statistic is small, it suggests that the GY series is probably well described as a DS-break A model.

It is interesting that the ***minimum*** value of the Perron-ADF statistic in Fig. 12 is the –4.99 value in 1972. The ZA/BLS method for selecting the break date would, therefore, have chosen 1972 not 1921 as the break date. Given the value of the test statistic, one would fail to reject the unit root with break hypothesis at the one or five percent significance levels; the respective critical values are –5.57 and –5.30 (assuming that the ZA asymptotic critical values still apply when a spike dummy is included in the ZAP-ADF equation as we do here). From Fig.10 showing the sequence of Wald statistics, on the other hand, it appears that the argument that the break occurs in 1972 rather than 1921 is weak.

Comparing our algorithm to the ZA/BLS algorithm suggests that the latter gives very little weight to the significance of the spike dummy. In effect, this amount to not taking seriously DS (unit root) with a type A break model. We believe this biases the results against the unit root hypothesis. Our algorithm should dominate the ZA/BLS algorithm in the situations described by LMN (1998). They emphasize the need to allow for the break under both the null and alternative hypotheses. We add to this point by stressing the need to apply an appropriate search algorithm for determining the break point.

The ZAP-ADF tests conducted here consider up to two break dates in the GY series. We tentatively conclude that the series is well characterized as a unit root process with a single level-shift break (type A) in 1921. Unfortunately, unit root tests have notoriously low power, so the common failure to reject the unit root hypothesis hardly provides a definitive determination of the true data generating process. An alternative approach is to consider the KPSS tests, which take stationarity or trend stationarity, rather than nonstationarity, as the null hypothesis. These tests are found in the appendix.

Two other arguments can also be invoked in making a choice between the TS and DS specifications:

- Plosser and Schwert (1978) discuss the pros and cons of estimating economic time series regression models, of which log-linear time trend models are a special case, is levels or first-differences. More precisely, they consider the relative costs of over-differencing and under-differencing. Which strategy is riskier: first-differencing (1)-(2), so that the DS model in (3)-(4) is estimated, when in fact there is no unit root in (1)-(2), or estimating (1)-(2) when there is, in fact, a unit root? They argue that "the problem of nonstationary disturbances (possibly in the levels regressions) are far more serious than the problems caused by excessive differencing (in the second differences regression, for example)." (1978, p.657).

Parameter instability in the TS model may, in fact, be an indication that the error process, in fact, has a unit root. Thus, we should look carefully for differences in the degree of parameter instability across the TS and DS specifications.

Given the uncertainty surrounding the question of unit roots, it seems reasonable to estimate both TS and DS models with one or two breaks. (The models with no breaks have already been estimated above.) Begin with the more general two-break specification.

iii. Estimated TS and DS Models with Two Breaks

Below we will consider the TS and DS model in turn, using our search algorithm to choose the dating of two break points.[27] As discussed above, we need to include only the level-shift and time interaction dummies to allow for breaks of type A, B, and C in the TS model. Thus the criterion for choosing the break dates (TB1,TB2) is the sup$\chi^2(4)$ statistic from the set of all $\chi^2(4)$ statistics testing the joint significance of the two dummies associated with all possible pairs of break dates. Analogously, in the DS specification, we need to include only the spike and level-shift dummies. The criterion is again a sup$\chi^2(4)$ statistic.

Once the two most plausible break points have been identified in the TS and DS specifications respectively, there are three subsamples of the GY index to consider. There is a necessary to estimate the growth rates for each segment: pre-TB1, TB1 through TB2, post-TB2. Estimates of the trend segments for both the TS and DS specifications are shown in Table 5. Also reported is the Wald test of the hypothesis that each trend

[27] Each specification requires the inclusion of two dummies for each break date. It can be shown that the break dates must be separated by at least one period to avoid perfect multicollinearity.

coefficient is equal to zero. A rejection of the hypothesis indicates the presence of a significant trend in the respective sub-period.

Table 5. Grid Search Results for Two Possible Breaks at Unknown Dates (TB1, TB2)[28]

Type of Model Type of Structural Break Dummies	**TS Model** Level & Time Interaction	**DS Model** Level & Spike
Chosen Break Points TB1 & TB2	**1921 & 1985**	**1921 & 1974**
Sup$\chi^2(4)$	34.43	47.40
Mean$\chi^2(4)$	8.07	3.35
(Segmented) Trend[1]		
1. pre_TB1	0.0032 (0.1836)	0.0027 (0.6559)
2. TB1 through TB2	-0.0006 (0.1298)	0.0001 (0.9690)
3. post_TB2	-0.0021 (0.5874)	-0.0109 (0.0307)
$\chi2$ stat(2)_TB1	13.25	19.32
$\chi2$ stat(2)_TB2	14.36	4.77

Note:
1. The p-value for the hypothesis that the trend coefficient is equal to zero is given in parenthesis. P values that are higher than your chosen test size (say .05) indicate failure to reject the null hypothesis of a zero trend for the given segment of the data. These p values ignore the fact that TB1 and TB2 were chosen so as to maximize sup$\chi^2(6)$. Thus the p-values on the trend segments are possibly inaccurate.

Examining the table, we find that sup $\chi^2(4)$ statistics for both the TS and DS specifications are 'large" (relative to the standard 1% critical value for $\chi^2(4)$ of 13.28. The mean $\chi^2(4)$ statistic for the DS model is very small, suggesting no issue of general parameter instability. The mean $\chi^2(4)$ statistic for the TS model is close enough to the standard critical vale that it is impossible to guess the outcome of a formal parameter stability tests based on simulated critical values.

The TS model estimation places the two breaks in 1921 and 1985. Moreover, the $\chi^2(2)$_TB1 and $\chi^2(2)$_TB2 stats for 1921 and 1985, respectively, are similar in magnitude, with 1985 being slightly larger (14.36 vs. 13.25, whereas the 1.0% critical value for

[28] In a Pentium III processor, the program runs for approximately 20 minutes each for the TS and DS models. In both cases, the maximum number of lags of the dependent variable considered (k) was six.

$\chi 2(2)$=9.21.) To calculate the segment-specific growth rates in the TS model, the formulas in the accompanying box are used.

Calculating Segment-Specific Growth Rates in the TS and DS Models

TS Model:

$$\ln(y_t) = \beta_0 + \beta_1 t + \beta_2 DUM_{TB1} + \beta_3 t * DUM_{TB1} + \beta_4 DUM_{TB2} + \beta_5 t * DUM_{TB2} + \sum_{i=1}^{k} \delta_i \ln(y_{t-i})$$

DS Model:

$$\Delta \ln(y_t) = \beta_0 + \beta_1 DUM_{TB1} + \beta_2 D(DUM_{TB1}) + \beta_3 DUM_{TB2} + \beta_4 D(DUM_{TB2}) + \sum_{i=1}^{k} \delta_i \Delta \ln(y_{t-i})$$

Segment-Specific Growth Rates in the TS Model:

Pre_TB1 growth rate: $\dfrac{\beta_1}{1-(\delta_1 + \ldots + \delta_k)}$

TB1 through TB2 growth rate: $\dfrac{\beta_1 + \beta_3}{1-(\delta_1 + \ldots + \delta_k)}$

Post_TB2 growth rate: $\dfrac{\beta_1 + \beta_3 + \beta_5}{1-(\delta_1 + \ldots + \delta_k)}$

Segment-Specific Growth Rates in the DS Model:

Pre_TB1 growth rate: $\dfrac{\beta_0}{1-(\delta_1 + \ldots + \delta_k)}$

TB1 through TB2 growth rate: $\dfrac{\beta_0 + \beta_1}{1-(\delta_1 + \ldots + \delta_k)}$

Post_TB2 growth rate: $\dfrac{\beta_0 + \beta_1 + \beta_3}{1-(\delta_1 + \ldots + \delta_k)}$

The resulting calculations for the TS model growth rates and their $\chi 2$ statistics (conventional p values noted) indicate that the trend in all three sub-periods are not statistically different from zero. In conclusion, therefore, if one rejects the unit root hypothesis and accepts the TS model, the GY series is best characterized as a zero-growth series that has experienced two significant downward level shifts (type A breaks), first in 1921 and then again in 1985.

Dependent Variable: GY
Method: Least Squares
Date: 10/31/01 Time: 11:25
Sample(adjusted): 1902 1998
Included observations: 97 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GY(-1)	0.621716	0.097600	6.370075	0.0000
GY(-2)	-0.314414	0.095788	-3.282405	0.0015
C	1.489242	0.202636	7.349340	0.0000
@TREND	0.002224	0.001737	1.280643	0.2036
DUM1921	-0.068788	0.025894	-2.656577	0.0094
DUM1921*@TREND	-0.002631	0.001782	-1.476821	0.1433
DUM1985	-0.012324	0.244391	-0.050425	0.9599
DUM1985*@TREND	-0.001040	0.002699	-0.385271	0.7010

R-squared	0.880814	Mean dependent var	2.026701
Adjusted R-squared	0.871440	S.D. dependent var	0.110272
S.E. of regression	0.039538	Akaike info criterion	-3.544222
Sum squared resid	0.139131	Schwarz criterion	-3.331875
Log likelihood	179.8948	F-statistic	93.96148
Durbin-Watson stat	1.840466	Prob(F-statistic)	0.000000

Fig. 13

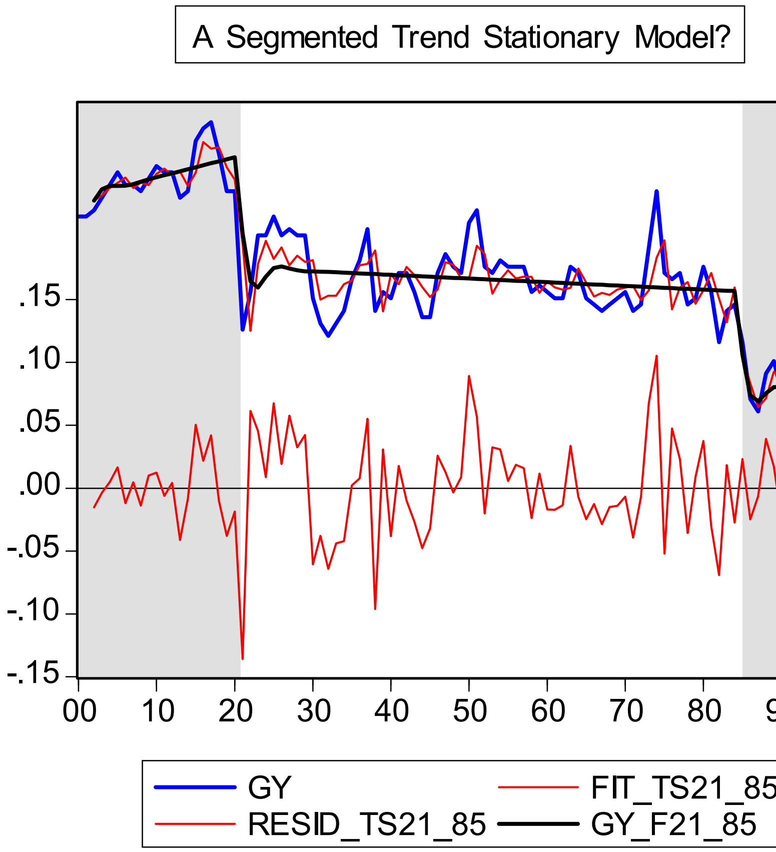


Figure 13 shows the actual logged GY series, the fitted values and residuals from the best fitting TS specification with two breaks, and the forecasted values starting in 1900 in order to show the long-run trend segments more clearly. The tests summarized in Table 2 above indicate that the trend is insignificantly different from zero in each of the three segments of the TS model: pre-1920, 1921-1984, and post-1984.

In contrast to the TS model, the DS model identifies the two break years as 1921 and 1974, rather than 1985. Note that for the DS model, the sup$\chi 2(4)$ is very large while the mean$\chi 2(4)$ statistic is quite small. (For comparison, the standard $\chi 2(4)$=13.28.) Also, the 1921 break has a much higher $\chi2(2)$ stat than the 1974 break. Together these $\chi2$ statistics suggest that, if one uses the DS specification, the GY series is well characterized by one (1921) or possibly two (1921, 1974) structural breaks rather than general parameter instability. Examining the $\chi^2(2)$_TB1 (=19.32) and the $\chi^2(2)$_TB2 (=4.77) statistics, it is clear that the 1921 break is significant, while the 1974 break is not statistically significant.[29] Thus the DS specification requires only a single break in 1921. This is consistent with our ZAP-ADF tests, which found a single break and were unable to reject the null hypothesis of a unit root.

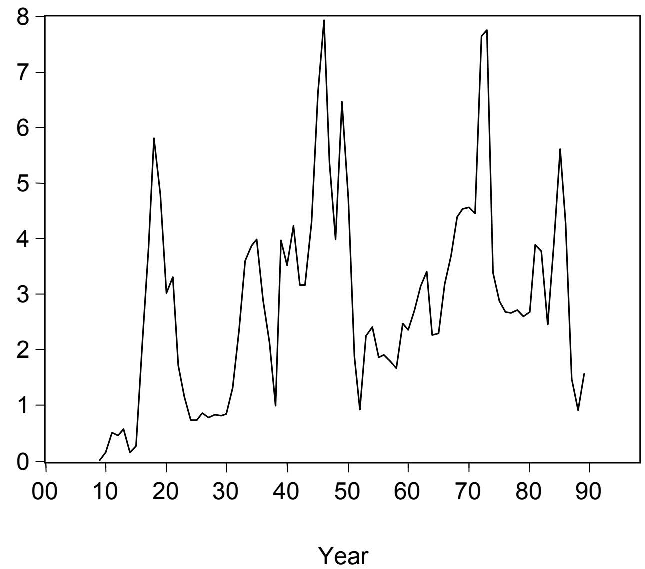
iv. Estimated TS and DS Models with a Single Break

We now estimate DS and TS Models with single breaks at an unknown date. We first search for one endogenous break in the GY series using the TS model. As one may recall, we need to include only the level dummy and the time interaction dummy in this particular setup. We obtain the sup$\chi^2(2)$ statistic that tests the hypothesis that these two dummies are equal to zero and graph it below. The maximum sup$\chi^2(2)$ has a value of 7.93 and occurs in 1946. The 1% critical value for the standard $\chi^2(2)$ distribution, however, is 9.21. Thus the sup$\chi^2(2)$ and mean$\chi^2(2)$ (=2.88) suggest that a TS model with zero breaks is adequate! Thus, rather curiously, the two-break model suggested that there are two (marginally?) significant breaks in 1921 and 1985, while the one break model finds no break at all! If one believes the two break model, the GY series has two downward level

[29] What about the calculated growth rates for each segment in the DS specification if we assume there are TWO breaks? Results for the DS model are slightly different from those obtained from the TS model. In spite of a statistically insignificant trend in each of the first two sub-periods, the DS model identifies the existence of a "possibly significant[29]" *negative* trend of 1.09% in the post-1974 period.

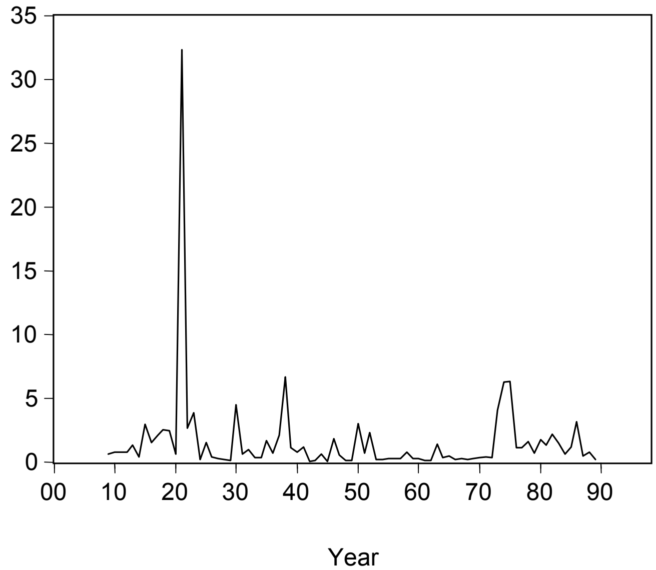
shifts, but not ongoing secular trend. If one believes the TS model with no breaks, there is a statistically significant negative time trend!

Fig. 14 $\chi 2(2)$ stats TS Model with One Endogenous Break



We now search for a single break in the GY series using the DS model. In this case, we include only the level and spike dummies in the estimation. We now use the sup$\chi 2(2)$ statistic to test the hypothesis that these two dummies are zero. Fig. 15 graphs the $\chi 2(2)$ for the DS model.

Fig. 15 The $\chi^2(2)$ Sequence for DS Model with One Break



Here, the maximum supχ2(2) has a value of 32.26 and occurs in 1921. The second highest supχ2(2) has a value of 6.28 and occurs in 1975. In addition, the mean$\chi^2(2)$ statistic is 1.58, a contrastingly low value compared to either the supχ^2 or the 1% critical value of 9.21 from the standard $\chi^2(2)$ distribution. Therefore, with the DS specification, a single downward level shift in 1921 but with no ongoing (stochastic) trend fits the data well.

6. CONCLUSIONS

Despite 50 years of empirical testing of the Prebisch-Singer hypothesis, a long-run downward trend in real commodity prices remains elusive. Previous studies have generated a range of conclusions, due in part to differences in data but mainly due to differences in specification, as to the stationarity of the error process and the number, timing, and nature of structural breaks. In this paper, we have attempted to allow the data to tell us the proper specification. In our most general specification (model 8, in Fig. 7), which allows for a unit root, and searches for two structural breaks of any kind, we find the most likely pair of breaks to be in 1921 and 1974, but 1974 break is statistically insignificant. Moreover, we cannot reject the hypothesis of a unit root. If we search for only one structural break, we find one very clearly in 1921, again with no rejection the unit root hypothesis. This model indicates also that there is no drift, either positive or negative, before or after 1921.

If we force the model to be trend stationary, we find much fuzzier results. The two-break model (model 7) puts the breaks in 1921 and 1985, with both breaks border-line significant. The three segments in this case (before, between and after the breaks), show no trend. The model with one break, puts the break in 1946, but is rejected in favor of model 1 (TS with no break). Only in the case of model 1, the model studied by researchers since the beginning of Prebisch-Singer testing, can one find a significant negative trend. Yet model 1 is inconsistent with our results N-step ahead forecasting.

We conclude the preponderance of evidence suggests that the series is well characterized as a unit root process with a single level-shift break (type A) in 1921.

7. **REFERENCES**

Andrews, Donald W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61(4), 821-856.

Ardeni, Pier Giorgio, and Brian Wright (1992), "The Prebisch-Singer Hypothesis: A Reappraisal Independent of Stationarity Hypothesis," *The Economic Journal*, Vol. 102, Issue 413, 803-812.

Banerjee, A., R. L. Lumsdaine, and J. H. Stock (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10(3), 271-287.

Bleaney and Greenaway (1993), "Long-run Trends in the Relative Price of Primary Commodities and in the Terms of Trade of Developing Countries," *Oxford Economic Papers* 45.

Bloch, H. and D. Sapsford (1997), "Some Estimates of the Prebisch and Singer Effects on the Terms of Trade between Primary Producers and Manufactures," *World Development*, Vol. 25, No. 11, 1873-1884.

Bloch, H. and D. Sapsford (2000), "Whither the terms of trade? An elaboration of the Prebisch-Singer hypothesis," *Cambridge Journal of Economics*, 24, 461-481.

Borensztein, E. and C. M. Reinhart (1994), "The Macroeconomic Determinants of Commodity Prices," *IMF Staff Papers*, Vol. 41, No. 2, 236-261.

Cashin, Paul, Hong Liang, and C. John McDermott (1999), "How Persistent are Shocks to World Commodity Prices?" IMF Working Paper.

Cashin, Paul and C. John McDermott (2001), "The Long-Run Behavior of Commodity Prices: Small Trends and Big Variability," IMF Working Paper.

Chambers, Marcus J., and Roy E. Bailey (1996), "A Theory of Commodity Price Fluctuations," *Journal of Political Economy*, vol. 104, No. 5, 924-957.

Christiano, Lawrence J. (1992), "Searching for a Break in GNP," *Journal of Business and Economic Statistics*, 10(3), 237-250.

Cuddington, John T. (1992), "Long-Run Trends in 26 Primary Commodity Prices: A Disaggregated Look at the Prebisch-Singer Hypothesis," *Journal of Development Economics*, 39, 207-227.

Cuddington and Urzúa (1989), "Trends and Cycles in the Net Barter Terms of Trade: A New Approach," *The Economic Journal*, Vol. 99, No. 396, 426-442.

Cuddington, John T., and H. Wei (1992), "An Empirical Analysis of Real Commodity Price Trends: Aggregation, Model Selection and Implications," *Estudios Economicos*, 7(2), 159-179.

Deaton, Angus, and Guy Laroque (1992), "On the Behavior of Commodity Prices," *Review of Economic Studies*, 59, 1-23.

Diakosavvas, Dimitis, and Pasquale L. Scandizzo (1991), "Trends in the Terms of Trade of Primary Commodities, 1900-1982: The Controversy and Its Origins," *Economic Development and Cultural Change*, 231-264.

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427-431.

Dixit, Avinash (1984), "Growth and the Terms of Trade under Imperfect Competition," in Kierzkowski, H., *Monopolistic Competition in International Trade*.

Enders, Walter (1995), *Applied Econometric Time Series.* New York: John Wiley and Sons.

Granger and Newbold (1974), " Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111-120.

Grilli, Enzo R., and M. C. Yang (1988), "Primary Commodity Prices, Manufactured Goods Prices, and the Terms of Trade of Developing Countries: What the Long Run Shows," *The World Bank Economic Review*, 2(1), 1-47.

Hadass, Yael and Jeffrey Williamson (2001), "Terms of Trade Shocks and Economic Performance 1870-1940: Prebisch and Singer Revisited," NBER Working Paper 8188.

Hansen, Bruce E. (1992), "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business & Economic Statistics*, 10(3), 321-335.

Hua, Ping (1998), "On Primary Commodity Prices: The Impact of Macroeconomic /Monetary Shocks," *Journal of Policy Modeling* 20(6), 767-790.

León, Javier and Raimundo Soto (1997), "Structural Breaks and Long-run Trends in Commodity Prices," *Journal of International Development* 3, 44-57.

Leybourne,Stephen J., Terence C. Mills, and Paul Newbold (1998), "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break Under the Null," *Journal of Econometrics* 87, 191-203.

Mehl, Arnaud (2000), "Unit Root Tests with Double Trend Breaks and the 1990s

Recession in Japan," *Japan and the World Economy*, 12, 363-379.

Nelson, C.R. and H. Kang (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica* 49, 741-751.

Perron, Pierre (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica* 57(6), 1361-1401.

Perron, Pierre (1993), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis: Erratum," *Econometrica* 61, 2248-249.

Perron, Pierre (1994), "Trend, Unit Root, and Structural Change in Macroeconomic Time Series," in B.B. Rao (ed.) *Cointegration for the Applied Economist*. New York, N.Y: Macmillan, pp. 113-146.

Perron, Pierre, and T. J. Vogelsang (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business & Economic Statistics*, 10(3), 301-320.

Phillips, Peter C.B. and Pierre Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika* 75, 335-346.

Ploberger, Werner, W. Kramer, and K. Kontrus (1989), "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307-318.

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

Powell, A. (1991), "Commodity and Developing Countries Terms of Trade: What Does the Long-Run Show?" *The Economic Journal*, 101, 1485-1496.

Prebisch, Raúl (1950), "The Economic Development of Latin America and its Principal Problems, reprinted in *Economic Bulletin for Latin America*, Vol. 7, No. 1, 1962, 1-22.

Reinhart, Carmen M. and Peter Wickham (1994), "Commodity Prices: Cyclical Weakness or Secular Decline?" *IMF Staff Papers*, Vol. 41, No. 2, 175-213.

Sapsford, D. (1985), "The Statistical Debate on the Net Barter Terms of Trade Between Primary Commodities and Manufactures: A Comment and Some Additional Evidence," *Economic Journal*, 95, 781-788.

Singer, H. W. (1950), "U.S. Foreign Investment in Underdeveloped Areas: The Distribution of Gains Between Investing and Borrowing Countries," *American Economic Review, Papers and Proceedings*, 40, 473-485.

Spraos, John (1980), "The Statistical Debate on the Net Barter Terms of Trade Between Primary Commodities and Manufactures," *The Economic Journal*, Vol. 90, No. 357, 107-128.

Zivot, Eric, and D. W. K. Andrews (1992), "Further Evidence on the Great Crash, the Oil - Price Shock, and the Unit Root Hypothesis," *Journal of Business & EconomicStudies*, 10(3), 251-270.

APPENDIX: KPSS Tests for Level and Trend Stationarity

Kwiatkowski, Phillips, Schmidt and Shin [KPSS] (1992) propose a test that examines the null hypothesis of stationarity or trend stationarity against the alternative of a unit root for a given series. To the present authors' knowledge, the KPSS test has not yet be employed in the PS literature.

The KPSS test involves decomposing the series into three components – a deterministic trend, a random walk, and a stationary error:

$$y_t = \xi t + r_t + \varepsilon_t \tag{9}$$

$$r_t = r_{t-1} + u_t \tag{10}$$

where r_t is a random walk, and u_t is an error process that is iid $(0, \sigma_u^2)$. The initial value of r_t is assumed to be fixed at r_0. The stationarity hypothesis consists of both trend stationarity and level stationarity. For instance, y_t is trend stationary under the null if σ_u^2 is equal to zero. In the special case where $\xi = 0$, the null hypothesis reflects stationarity around a level r_0. The authors derive the test statistic for the trend stationary case, $\hat{\eta}_\tau$, by obtaining the partial sum process of the residuals from (9) and modeling it as a Brownian motion. The test statistic for the level stationary case, $\hat{\eta}_\mu$, is derived in exactly the same manner except that residuals from (9) are now obtained from a regression of y_t on an intercept only rather than on both the intercept and the trend.. Here, we apply the KPSS test to the real GY index and obtain the following results:[30]

	Lag Truncation Parameter										
Test Statistic	0	1	2	3	4	5	6	7	8	9	10
$\hat{\eta}_{\tau}$	0.40	0.23	0.18	0.15	0.13	0.12	0.11	0.10	0.10	0.09	0.09
$\hat{\eta}_{\mu}$	5.67	3.02	2.11	1.65	1.37	1.17	1.04	0.93	0.85	0.78	0.73

Table 6. Stationarity Test Results for GY series

Note: Critical Values for $\hat{\eta}_{\tau}$ and $\hat{\eta}_{\mu}$ presented by Kwiatkowski et al. (1992)

$\hat{\eta}_{\tau}$				
Critical Level	**0.10**	**0.05**	**0.025**	**0.001**
Critical Value	0.119	0.146	0.176	0.216
$\hat{\eta}_{\mu}$				
Critical Level	**0.10**	**0.05**	**0.025**	**0.001**
Critical Value	0.347	0.463	0.574	0.739

For the lag truncation parameter two and above, we cannot reject the hypothesis of trend stationarity even at the 1% significance level. On the other hand, we can reject the hypothesis of level stationarity at high significance levels for all lags considered.[31] Based on the KPSS test, therefore, the real GY series appears to be stationary around a deterministic trend. Thus, the results of the ZAP-ADF tests and KPSS tests are

[30] KPSS applied their new test to the Nelson-Plosser data and find that the hypothesis of trend stationarity cannot be rejected for many series while that of level stationarity can be rejected for most of the series.
[31] Kwiatkowski et al. (1992) state that the ability to reject the hypothesis of level stationarity in their series is not very surprising due to the obvious deterministic trends present in the series.

inconsistent, with the former pointing toward the DS model and the later favoring the TS specification.