Unit Roots, TV–STARs, and the Commodity Terms of Trade: A Further Assessment of the Prebisch–Singer Hypothesis*

Selected Paper prepared for presentation at
the American Agricultural Economics Association Annual Meeting
Long Beach, CA, July 23-26, 2006

Joseph V. Balagtas†
Department of Agricultural Economics
Purdue University

Matthew T. Holt‡
Department of Agricultural Economics
Purdue University

This Draft: June 27, 2006

Abstract

This paper examines the Prebisch–Singer hypothesis, that is, the hypothesis of a long-run decline in the prices of primary commodity prices relative to that of manufactured goods, by considering nonlinear alternatives. Specifically, in this paper we use bootstrap procedures to test the linear unit root model against models belonging to the family of smooth transition autoregressions (STARs) for 24 commodities, 1900–98. In eighteen cases we reject the linear null at usual significance levels. In fourteen cases we are able to successfully fit STAR-type models. Simulation results show there is very little support for the Prebisch–Singer hypothesis.

Keywords: Nonlinear model; Primary commodities; Smooth Transition autoregression; Time-varying autoregression; Unit root tests

JEL Classification Codes: O13; C12; C22; C52

*This research has been supported by the Agricultural Experiment Station at Purdue University. Senior authorship is not assigned.

†Corresponding author: Department of Agricultural Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907–2056, USA. Telephone: 765-494-4298. Fax: 765-494-9176. E-mail: balagtas@purdue.edu.

‡Department of Agricultural Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907–2056, USA. Telephone: 765-494-7709. Fax: 765-494-9176. E-mail: mholt@purdue.edu.

Introduction

An issue of continuing interest in international economics is the so called Prebisch-Singer, or PS, hypothesis (Prebisch, 1950; Singer, 1950). According to the PS hypothesis commodity prices decline over the long run relative to the prices of manufactured goods, thereby placing developing countries (i.e., primary producers and exporters of commodities) at a disadvantage relative to developed countries (i.e., producers and exporters of manufactured goods). In this manner there will be a long-term deterioration in the commodity terms of trade.

Of course the validity of the PS hypothesis is, at root, an empirical question, and one which has received considerable attention over the years. See, for example, Grilli and Yang (1988), Cuddington and Urzua (1989), Powell (1991), and Ardeni and Wright (1992). In modern, time series terms, a key question surrounding the PS hypothesis is whether the relative commodity price series in question contains a deterministic trend or is associated with a unit root (i.e., stochastic trend). The original work on the PS hypothesis assumed that commodity prices were trend stationary (Prebisch (1950), Singer 1950). In more recent work researchers have tested for and, if called for, incorporated stochastic trends (i.e., difference stationarity). Examples of studies of this sort include Newobold and Vougas (1996) and Kim et al. (2003). In the later case there is generally less empirical support for long-term attenuation in the commodity terms of trade. The results may also be sensitive to the period used.

A key issue in testing the PS hypothesis is whether the underlying price series has experienced structural change, especially since relatively long time periods are preferably employed in any empirical analysis. Ocampo and Parra (2004) argued that deteriorations in the terms of trade have been discontinuous, with the 1920s and the 1980s being periods for which the decline in the commodity terms of trade was particularly notable. To the extent that structural breaks have been observed in the data, standard unit root tests may provide misleading results (Perron, 1989). For these reasons, recent research examining the PS hypothesis has focused on employing unit root tests where the possibility of structural breaks is allowed. Relevant examples include Leon and Soto (1997), Zanias (2005), and Keller and Wohar (2006). The empirical results generally support the observation that, in contrast to the PS hypothesis, deteriorations in the terms of trade have been discontinuous and episodic.

While considerable progress has been made in terms of the econometric framework employed in examining the PS hypothesis, more work is required. To begin, research on this topic has yet to consider the question of whether or not commodity price dynamics are also changing with time. Specifically, previous research has considered only the case where the intercept term and/or the coefficient associated with a linear trend term have experienced structural change. A more complete analysis would also examine the possibility for structural change in the model's autoregressive coefficients. Alternatively, the data may also exhibit nonlinear features with, for example, small shocks having potentially different impacts from large shocks. This case was considered explicitly by Persson and Tersvirta (2003), who used the Grilli and Yang (1988) data to estimate a member of the family of smooth transition autoregressions (STARs). These authors did not, however, formally test for nonstationarity, an important component of any formal assessment of the PS hypothesis. Finally, it is possible that the underlying commodity price data exhibit features consistent with both structural change and nonlinearity. If so, it is important that each of these aspects be incorporated in any formal testing framework, including tests of the unit root hypothesis.

In this paper we investigate whether the commodity price data exhibit unit roots, either globally or possibly locally. The basic methodology to be used builds upon previous research by Tersvirta (1994) and will examine the potential for STARs, time–varying autoeregressions (TVARs) (Lin and Tersvirta (1994)), and models that contain both nonlinearity and time variation, or TV-STARs (Lundbergh et al. (2003)). Each of these models employs a member of the family of univariate logistic functions to capture structural change or nonlinearity. As well, a key feature of the TVAR is that by including higher-order trend terms in the logistic function, structural change may be nonmonotonic and, depending on the speed of adjustment, instantaneous, therefore implying that multiple structural breaks (changes) may occur. This later feature may, of course, be important for investigating the PS hypothesis.

This study uses the data on individual commodity prices, 1900–98, that represents an extension of the original Grilli and Yang (1998) index, and used recently by Kim et al. (2003) and Kellard and Wohar (2006). Building on Eklund (2003), we test a linear unit root model against STAR–type alternatives. A low-order Taylor series expansion of the underlying STAR–type model is used as an approximation, therefore eliminating unidentified nuisance parameters. This model is then used as the alternative in a series of nonparametric bootstrap simulations wherein the unit root hypothesis is formally tested. As well, Tersvirta's (1994) testing framework will be used to determine if, in fact, the data are adequately characterized by a linear model, or whether a STAR-type or TVAR–type specification might be required. If a nonlinear model is called for this model will be econometrically estimated, with the resulting model subjected to a battery of diagnostic tests. Among other things, the resulting model will be used to recover any trends in the commodity terms of trade over time. Finally, if nonlinear features are identified, the implications of the PS hypothesis are examined by using stochastic simulations along forward paths.

A key result is that for eighteen of the 24 commodities examined, the linear unit–root model is rejected in favor of either a STAR–type alternative. Moreover, we are able to successfully fit a STAR–type model to the data in fourteen instances. A central finding, and one that is largely consistent with more recent research in the area, is that while nonlinearity and parameter nonconstancy often are relevant features of the model, in no case do we find substantial evidence in support of the PS hypothesis.

The Modeling Framework

As noted in the introduction, a central thrust of the present paper is to model appropriately identified commodity prices by using nonlinear time series techniques, specifically, by using one or more members of the family of smooth transition autoregressions (STARs). The staring point in any STAR–type modeling exercise is the linear autoregressive (AR) model. Let y_t denote the (natural logarithm) of a commodity price. A corresponding pth–order AR model is then simply specified as

$$\Delta y_t \ = \ \alpha + \beta y_{t-1} + \boldsymbol{\phi}' \boldsymbol{x}_t + \varepsilon_t, \tag{1}$$

where Δ is a first difference operator such that $\Delta y_{t-k} = y_{t-k} - y_{t-k-1}$ for any $k > 0$, $\boldsymbol{x}_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})'$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ is a vector of autoregressive parameters to be estimated, and ε_t is an additive error process such that $\varepsilon_t \sim iid(0, \sigma^2)$. As written in (1) the AR model does not impose a (global) unit root; such a specification may, however, be obtained by simply imposing the restriction $\beta = 0$. Aside from determining whether or not a unit root should be imposed, the modeling exercise must also determine the lag order p for the model, perhaps by using data–based procedures such as the AIC or BIC. See, for example, Hall (1994).

Variants of the AR model in (1) have been used in recent years to model commodity price data, either individually or in aggregate, and to otherwise examine the PS hypothesis. Examples include Newbold and Vogas (1996), Leon and Soto (1997), and Kim et al. (2000). In the context of the unit root version of (1), a statistically significant estimate of the drift term α–and assuming that its sign is negative–is taken as *premia facie* evidence in favor of the PS hypothesis.

A potential limitation of testing for a unit root hypothesis in the context of (1) is that isolated structural breaks may bias the results in favor of finding that $\beta = 0$ (see, e.g., Perron, 1989). For example, suppose that the model in (1) is associated with a single, discrete structural break and trend break. Assume these breaks occur at time t_b, such that $0 < t_b < T$. We might then specify the model in (1) alternatively as

$$\Delta y_t \;=\; \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}' \boldsymbol{x}_t + (\alpha_1 + \beta_1 y_{t-1}) D_{b,t} + \varepsilon_t, \tag{2}$$

where $D_{b,t} = 1$ if $t > t_b$, and is 0 otherwise.[1] Tests of models similar (2) against a linear unit root model without trend breaks, that is, against equation (1) with $\beta = 0$, and where the breaks are determined as part of the testing/estimation framework, have been proposed by Bannerjee et al. (1992), Zivot and Andrews (1992), and Perron (1997), among others. Lumsdaine and Papell (1997) extend the framework to allow for two distinct shifts in intercept and trend terms. These methods have been applied recently by Zanias (2005) and Kellard and Wohar (2006) in testing for unit roots in, respectively, an aggregate commodity price index and a series of 24 individual commodity prices. Although the results are somewhat mixed, there is considerable evidence that many commodity prices are stationary once structural breaks are considered.

[1] A more typical specification for (2) is $\Delta y_t \;=\; \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}' \boldsymbol{x}_t + \widetilde{\alpha}_1 D1_{b,t} + \widetilde{\beta}_1 D2_{b,t} y_{t-1} + \varepsilon_t$, where $D1_{b,t}$ is identical to $D_{b,t}$ above, and $D2_{b,t} = (t - t_b) D1_{b,t}$. These two specifications are essentially identical in that the model in (2) simply absorbs the term $-t_b D1_{b,t}$ into the β_1 parameter, that is, $\beta_1 + t_b = \widetilde{\beta}$.

Moreover, once structural breaks are allowed for there is generally less evidence in support of the PS hypothesis (Kellar and Wohar, 2006).

While allowing for one or possibly two breaks in intercept and trend terms admits richer alternatives in testing the unit root hypothesis–and, by extension, the PS hypothesis–the above does not exhaust the full range of alternatives. For example, none of these tests allow the model's higher-order dynamics, that is, the autocorrelation parameters in the vector ϕ, to change. Moreover, prior research has not considered the possibility that structural change is a potentially smooth process over time. Finally, with the exception of Persson and Teräsvirta (2003), prior research has not investigated the possibility that nonlinearity could be a feature of historical commodity price data. To begin, consider the following generalization of (1)

$$\Delta y_t = \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{'} \boldsymbol{x}_t + (\alpha_1 + \beta_1 y_{t-1} + \boldsymbol{\phi}_1^{'} \boldsymbol{x}_t) G(s_t; \gamma, c) + \varepsilon_t, \tag{3}$$

where $s_t = t^* = t/T$. In (3) $G(t^*; \gamma, c)$ is the so–called transition function, which in the spirit of the structural break model in (2) has a value bounded between 0 and 1. The main difference, however, is that intermediate values, that is, values within the unit interval, are now admitted, and thus $D_{b,t}$ is no longer restricted to be a Heaviside indicator function. Moreover, unlike (2) the specification in (3) allows the autocorrelation coefficients, the ϕ_i's, to vary over time. As written, the model in (3) is a member of the family of time–varying autoregressions, or TVARs, introduced initially by Lin and Teräsvirta (1994). Note in particular that if $\beta_0 = \beta_1$ and $\boldsymbol{\phi}_0 = \boldsymbol{\phi}_1$, the model in (3) would provide a direct way of incorporating the PS hypothesis as originally envisaged by Prebisch (1950) and Singer (1950).

Of course it is possible to generalize the model in (3) in one or more substantive ways. One important generalization is to note that regime change, that is, structural change, need not be triggered solely or in fact, at all, by a trend variable. Specifically, if t^* is replaced in $G(.)$ by a variable that is a continuous function of the lagged endogenous variable $y_{t-d}, d > 0$, say, $s_t = f(y_{t-d})$, then (3) becomes a member of the family of smooth transition autoregressions, or STARs, as described initially by Teräsvirta (1994). Here s_t is refereed to as the transition variable and d as the delay parameter. Henceforth we adopt the following notation. The variable

s_t will be used to denote the transition variable, and unless otherwise indicated it will be used generically to signify either t^* or $f(y_{t-d})$.

A central question then is how exactly to specify the transition function in (3). Several alternatives exist, including the first–order logistic and the exponential functions. These are specified respectively as

$$G(s_t;\gamma,c) \;=\; (1+\exp\{-\gamma(s_t-c)\})^{-1},\;\gamma>0 \tag{4}$$

and

$$G(s_t;\gamma,c) \;=\; 1-\exp\{-\gamma(s_t-c)^2\}),\;\gamma>0. \tag{5}$$

If (4) is used in conjunction with (3), the resulting model is a member of Teräsvirta's (1994) logistic smooth transition autoregression (LSTAR) family. Alternatively, if (5) is used in conjunction with (3), the resulting model is a member of Teräsvirta's (1994) exponential smooth transition autoregression (ESTAR) family.

For both transition functions γ is referred to as the speed of adjustment parameter. More specifically, the specification in (4) is such that as $\gamma\to\infty$ then $G(.)$ becomes a Heaviside indicator function, that is, a function such that $G(.)=0$ for $s_t<c$ and $G(.)=1$ for $s_t>c$. In this manner the model in (2) is a special case of the LSTAR. Alternatively, for (5) as $\gamma\to\infty$ or $\gamma\to-\infty$ then $G(.)\to 1$. In (4) c, the location parameter, indicates exactly the point where $G(c;\gamma,c)=0.5$, whereas in (5) $G(c;\gamma,c)=0$. Therefore, the structural break model in (2) is also a special case of the ESTAR. Finally, in the particular instance where $s_t=t^*$ and where (4), that is, in the case where a TVAR is specified, and when $\gamma\to\infty$, the resulting model becomes one of discrete structural change with a single break point.

A generalization of (4) is also available, a generalization that, in fact, may be quite useful in empirical work when structural change over time is being modeled. Specifically, the general nth–order logistic function is defined as

$$G(s_t;\gamma,c) \;=\; (1+\exp\{-\gamma\prod_{i=1}^{n}(s_t-c_i)\})^{-1},\quad \gamma>0, c_1\le c_2\le\ldots\le c_n, \tag{6}$$

where c is now a vector such that $c = (c_1, c_2, \ldots, c_n)$. If n is set to two in (6), the resulting model, called a quadratic STAR, or QSTAR, is similar in several respects to an ESTAR. For example, as $\gamma \to \infty$ and assuming $c_1 \neq c_2$, $G(s_t; \gamma, c)$ will equal unity for $s_t < c_1$ and for $s_t > c_2$ and zero in between. The central difference is that as $\gamma \to 0$ the underlying model becomes linear in parameters. Of course the ESTAR is slightly more parsimonious than the QSTAR. In the case where $n > 2$, considerable flexibility is afforded in the model's specification. In particular, and in the context of a TVAR, if $n \geq 3$ any structural change need no longer be monotonic function of t^*. As well, and in the spirit of the prior discussion on testing for unit roots against alternatives that are stationary but allow for one or more structural breaks, the transition function in (6) has enough flexibility to accommodate such a situation. See, for example, Lin and Teräsvirta (1994). To recapitulate, when (6) is combined with (3) and when $s_t = t^*$, the resulting model will readily nest the one– and two–regime change models considered respectively by Zanias (2005) and by Kellard and Wohar (2006).

A distinct possibility is that the model in (3) does not adequately capture all of the essential features of the data. Specifically, and in the spirit of single hidden layer feed–forward artificial neural networks (ANNs), the model in (3) may require additional (additive) nonlinear components. For example, a three–regime model may be specified as

$$\begin{aligned}\Delta y_t &= \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{'} \boldsymbol{x}_t + (\alpha_1 + \beta_1 y_{t-1} + \boldsymbol{\phi}_1^{'} \boldsymbol{x}_t) G_1(s_t; \gamma_1, c_1) \\ &+ (\alpha_2 + \beta_2 y_{t-1} + \boldsymbol{\phi}_2^{'} \boldsymbol{x}_t) G_2(s_t; \gamma_2, c_2) + \varepsilon_t. \end{aligned} \tag{7}$$

Additional details on this model specification are provided in Eitrheim and Teräsvirta (1996). Importantly, these authors also describe a framework in which Lagrange multiplier (LM) tests may be constructed to test the model in (3) for remaining additive nonlinearity of the sort implied in (7).[2] Of interest is that the specification in (7) may accommodate a situation in which both structural change (i.e., $s_t = t^*$) and nonlinearity (i.e., $s_t = f(y_{t-d})$ might be prevailing features of the data. That is, the model in (7) is a form of the TV–STAR models developed

[2]More elaborate versions of (7), referred to as multiple regime STARs, or MRSTARs, have been developed by van Dijk and Franses (1999). These models are not considered here, however, because substantially more data than are presently available for empirical tests of the PS hypothesis are required for their implementation.

originally by Lundbergh, Teräsvirta, and van Dijk (2003). With the exception of Holt and Craig (2006), models of this sort have generally not been considered in prior research on commodity price behavior.

The Testing Framework

In view of the foregoing discussion, several important questions remain. First, how is it possible to know if nonlinearity is truly a feature of the data, that is, how might linearity tests be performed? And second, and not unrelated to the first question, how might tests for (global) unit roots be performed when either nonlinearity or structural change is considered as an alternative? Both issues are now examined in greater detail.

Linearity Testing

Regarding the first question, it is desirable to have a method of testing the linear model in (1) against time–varying and/or nonlinear alternatives such as those in (3) or (7). At first blush this would seem to be a simple task in that a simple test of the statistical significance of the estimated γ parameter in the relevant transition function should surely suffice. Such an approach is not appropriate in the present context, however, because there are unidentified nuisance parameters under the null hypothesis, that is, the linearity hypothesis, notably the autoregressive coefficients implied in ϕ_1 and the constant term a_1. In the statistics literature this is generally referred to as the Davies (1978, 1988) problem, a problem that is, moreover, pervasive in the literature on testing linear models against nonlinear or threshold alternatives. See, for example, Andrews and Ploberger (1994) and Hansen (1996). The implication is that the sampling distribution for the estimator of γ no longer has the usual asymptotic properties, and therefore standard asymptotic tests (i.e., standard t–tests, etc.) no longer apply.

In the case where smooth transition models are considered as an alternative, Luukkonen, Saikkonen, and Teräsvirta (1988) have proposed one workable solution to the Davies–type problem. Specifically, they recommend replacing $G(s_t; \gamma, c)$ in (3) with a suitable Taylor series approximation. For example, if a third–order Taylor series in s_t is used to approximate $G(.)$, we may rewrite (3) as follows

$$\Delta y_t \;=\; \boldsymbol{\delta}_1^{'}\widetilde{\boldsymbol{x}}_t + \boldsymbol{\delta}_2^{'}\widetilde{\boldsymbol{x}}_t s_t + \boldsymbol{\delta}_3^{'}\widetilde{\boldsymbol{x}}_t s_t^2 + \boldsymbol{\delta}_4^{'}\widetilde{\boldsymbol{x}}_t s_t^3 + e_t, \tag{8}$$

where $\widetilde{\boldsymbol{x}}_t = (1, y_{t-1}, \boldsymbol{x}_t^{'})^{'}$ and where the parameters in (8) are functions of the original parameters in (3) including the speed–of–adjustment parameter γ and the location parameter c. Also, the error term e_t in (8) is a function of the original error term ε_t as well as a remainder term $R_3(s_t;\gamma,c)$. It follows that under the null hypothesis of linearity $R_3(s_t;\gamma,c) = 0$ and $e_t = \varepsilon_t$; standard testing strategies may therefore be applied.

As noted by van Dijk, Teräsvirta, and Franses (2003), it is straightforward to deduce that the parameter $\delta_{1,1}$ and the parameter vectors $\delta_i, i = 2,3,4$, are functions of the original parameters $a_0, a_1, \beta_0, \beta_1, \boldsymbol{\phi}_0$, and $\boldsymbol{\delta}_1$ in (3) such that testing $H_0^{''} : \boldsymbol{\delta}_2 = \boldsymbol{\delta}_3 = \boldsymbol{\delta}_4 = 0$ now constitutes a direct test of the linearity hypothesis $H_0^{'} : \gamma = 0$. This test, denoted as F_L, may be performed as a standard extra–sum–of–squares test, which under the null hypothesis will have a test statistic that is distributed asymptotically as an F statistic with $3(p+2)$ and $T-(p+2)$ degrees of freedom.[3] As outlined by Teräsvirta and Anderson (1992), Lin and Teräsvirta (1994), and Teräsvirta (1994), an additional testing sequence may be pursued in an attempt to identify whether an LSTAR or ESTAR model is more appropriate. Specifically, (8) may be used to test the following sequence of hypotheses:

$$H_{04}\text{: } \delta_{4j} = 0,\ j = 1,\ldots,p+2, \tag{9}$$

$$H_{03}\text{: } \delta_{3j} = 0 \,|\, \delta_{4j} = 0,\ j = 1,\ldots,p+2, \tag{10}$$

$$H_{03}\text{: } \delta_{2j} = 0 \,|\, \delta_{3j} = \delta_{4j} = 0,\ j = 1,\ldots,p+2. \tag{11}$$

F versions of the LM tests associated with testing the hypotheses in (9), (10), and (11) are referred to respectively as F_4, F_3, and F_2. Assuming that linearity is rejected, that is, that $H_0^{''}$ is rejected, then an LSTAR is chosen if either F_4 or F_2 has the minimal p–value in the sequence.

[3] Alternatively, and in the usual spirit of LM–type tests, a χ^2 version of the linearity test with $3(p+2)$ degrees of freedom could also be constructed. In small samples of the sort used here, however, the χ^2 test may be seriously oversized while the F version of the test generally has better size properties. For this reason we rely throughout on the corresponding F version of all LM–type tests.

Otherwise, an ESTAR is chosen. See Teräsvirta (1994) for additional details.

Combined Unit Root and Linearity Testing

Newbold and Vougas (1996) demonstrated that the evidence for the Prebisch-Singer hypothesis is weaker if the commodity price data are generated by a unit-root process. Thus, testing for the presence of unit roots is a crucial part of evaluating the PS hypothesis.

As noted previously, in the context of a linear autoregressive model standard unit root tests may be conducted by testing whether or not $\beta = 0$ in (1), assuming of course that structural change and/or nonlinearly are not a feature of the data. See Dickey and Fuller (1979, 1981) and Said and Dickey (1984). And as also observed before, numerous studies have investigated tests of linear unit root models against stationary alternatives that allow for one and possibly two structural breaks. However, testing the unit version of (1) against nonlinear alternatives such as those depicted in (3) is still an emerging field of inquiry. For example, Enders and Granger (1998) and Canner and Hansen (2001) test the unit root hypothesis against stationary threshold autoregressive (TAR) alternatives. Alternatively, Kapetanios, Shin, and Snell (2003)propose a test of the linear unit root model against an alternative that allows the constant term to change according to an ESTAR transition function. Finally, Eklund (2003) has recently proposed a test of the linear unit root model against a specific LSTAR alternative. Among other things, Eklund (2003) finds that bootstrapping the relevant test statistics enhances the size properties of the relevant tests in certain limiting cases. In what follows we adopt the overall strategy proposed by Eklund (2003), although with some modifications.

What is desired here is a test of the unit root version of the linear autoregression in (1) against the more general STAR–type alternative in (3). That is, we seek a test of (1)

$$\Delta y_t \;=\; \alpha + \boldsymbol{\phi}^{'} \boldsymbol{x}_t + \varepsilon_t, \tag{12}$$

against the alternative

$$\Delta y_t \;=\; \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{'} \boldsymbol{x}_t + (\alpha_1 + \beta_1 y_{t-1} + \boldsymbol{\phi}_1^{'} \boldsymbol{x}_t) G(s_t; \gamma, c) + \varepsilon_t, \tag{13}$$

where $\boldsymbol{x}_t$ and s_t are as previously defined and $G(.)$ is given by either (4) or (5). For reasons already mentioned, that is, for reasons associated with unidentified nuisance parameters under the null, it is not possible to directly test (12) against (13).

As discussed above, it may be possible to circumvent this problem by replacing $G(s_t; \gamma, c)$ with a suitable Taylor series approximation. For example, if, as before, $G(s_t; \gamma, c)$ is replaced by a third–order approximation, we obtain

$$\Delta y_t \;=\; \delta_0 + \lambda_0 y_{t-1} + \boldsymbol{\vartheta}_0^{'} \boldsymbol{x}_t + \sum_{i=1}^{3} \delta_i s_t^i + \sum_{i=1}^{3} \lambda_i y_{t-1} s_t^i + \sum_{i=1}^{3} \boldsymbol{\vartheta}_i^{'} \boldsymbol{x}_t s_t^i + \xi_t, \tag{14}$$

where, as before, ξ_t is a function of a remainder term $R_3(s_t; \gamma, c)$ as well as ε_t in (13). It is now possible to use the auxiliary regression in (14) to directly test for *both* linearity and a unit root in the underlying y_t series. Specifically, testing the hypothesis $H_0^{lur} : \lambda_0 = \delta_1 = \delta_2 = \delta_3 = \lambda_1 = \lambda_2 = \lambda_3 = \vartheta_{1,1} = \ldots = \vartheta_{3,p} = 0$ constitutes such a test inasmuch as the linear autoregression with a single unit root in (12) is obtained. Note also that under the null hypothesis that $\varepsilon_t = \xi_t$ since the remainder term R_3 is identically zero. It is, moreover, possible to construct the usual F test, denoted here by F_{lur}, associated with imposing H_0^{lur}. In general F_{lur} will be associated with $(7 + 3p)$ and $T - (8 + 4p)$ degrees of freedom.

The problem in the present case, is that for conventional reasons the standard F statistic is no longer associated with the usual limiting distribution under the null of linearity and a unit root. Eklund (2003) did, in fact, obtain asymptotic results for a special case of (14). Specifically, he worked with the alternative model

$$\Delta y_t \;=\; \delta_0 + \lambda_0 y_{t-1} + \vartheta_{0,1} \Delta y_{t-1} + \vartheta_{0,1} \Delta y_{t-1} s_t + \xi_t, \tag{15}$$

where, moreover, $s_t = y_{t-1}$.

While testing (12) against (15) allows for the development of asymptotic results, the tradeoff is that the alternative in (15) lacks considerable generality and flexibility. Alternatively, it is always possible, as in Li and Maddala (1996) and Enders and Granger (1998), and as suggested by MacKinnon (2002), to simply obtain the empirical distributions of the relevant F_{lur} statistic by using simulation methods. Of course this method has no particular grounding in asymptotic theory, and therefore it may be of some limited appeal, but it does allow for greater flexibility in modeling and testing than would otherwise be the case; we leave it to future research to derive asymptotic results for testing (12) against the more general model in (14).

In implementing our test of the nonstationary AR model in (12) against the (possibly locally) stationary nonlinear model in (15), we estimate both models using the observed sample data and obtain the sample estimate for F_{lur}. We then use a dynamic bootstrap of the null model's estimated residuals to construct a reasonably large number, B, of pseudo samples (Li and Maddala, 1996). Both the null and the auxiliary regression models are re–estimated on each pseudo data, and, as well, simulated values for the F_{lur} statistic are obtained. To construct an empirical p–value for the F_{lur} test statistic we simply observe the fraction of times that the sample value exceeds the corresponding simulated values. In the empirical implementation we base our results on $B = 999$ bootstrap replications.

Data

In the empirical analysis we use annual data on prices for 24 primary commodities spanning 1900–98. Nominal prices are deflated by the United Nations Manufactures Unit Value index. These data were recently used by Kellard and Wohar (2006) and Kim et al. (2003), and are also used to calculate the commodity price index developed by Grilli and Yang (1988), and later extended by the Cashin and McDermott (2002). We conduct our analysis on the natural logarithms of the various price series. Plots of the raw data are not reported to conserve space; they are available in Kellard and Wohar (2006).

Results

Linear Unit Root Tests

The analysis begins with an examination of the linear unit root hypothesis, that is, with tests for a single unit root against an alternative that is stationary in the levels. To this end we compute augmented Dickey–Fuller (ADF) tests for the case where the model includes an intercept but may or may not include a linear trend (i.e., the t_τ and t_μ test statistics, respectively). Specifically, models similar to (1) are estimated for each commodity with and without a trend term. In each case the AIC is used to choose the optimal lag length, where a maximum of six lags may be included. Approximate p–values are constructed by performing B=999 dynamic, non–parametric bootstrap simulations. The results are recorded in Table 1.

Based on the preliminary evidence in Table 1, there is substantial support for the unit root hypothesis. For example, in the case where the trend is excluded the null hypothesis is rejected at the 5% level for only tobacco and zinc. If a trend is included, the hypothesis is rejected at the 5% level for hides, rubber, and timber, and at the 10% level for wheat. While there is some discrepancy in results depending on whether a trend is included, the overall picture emerging from Table 2 is one of general support for the unit root hypothesis in commodity prices. This conclusion is, moreover, consistent with results reported elsewhere by Kim et al. (2003) and Kellard and Wohar (2006), among others.

Unit Roots versus Nonlinearity

Of course the ADF results reported in Table 1 do not take account of any structural change or nonlinearities, features that might otherwise be part of the data generating process. To examine this issue in greater detail, we perform tests of the linear unit root model against stationary nonlinear and time–varying alternatives by using the general testing framework outlined in the previous section. Specifically, for each commodity a model consistent with (12) is tested against a model consistent with (15), the point being to construct the F_{lur} statistic. Approximate p–values for this statistic are then constructed by using bootstrap procedures. Regarding tests for nonlinearity, we follow Persson and Teräsvirta (2003) in using $s_t = \Delta y_{t-d} = y_{t-1-d} - y_{t-2-d}$ as the candidate transition variable, where $d = 1, \ldots, 6$. As with the ADF tests, lag lengths are determined by using the AIC. Results are recorded in Table 2. With several exceptions, noted below, we report results associated with only the minimal p–value for the F_{lur} test statistic across candidate transition variables for all 24 commodities.

As reported in Table 2, the null model is rejected for 16 commodities at the 5% level including aluminum, beef, cocoa, copper, cotton, hides, lamb, lead, palmoil, rubber, silver, sugar, tea, timber, tin, and tobacco. An additional two, wool, and zinc, are associated with a rejection of the null at the 10% level. Only in the case of bananas, coffee, jute, maize, rice, and wheat does a linear unit root model seem to adequately characterize the data.[4] For four commodities (aluminum, cotton, silver, and sugar) the linear unit root model is rejected rather strongly for more than one candidate transition variable. We report these additional results because, as noted in additional detail below, some added degree of latitude might be afforded in specifying and estimating a STAR–type model for these commodities.

One surprising result revealed in Table 2 is that, with the exception of tin, all commodities for which the null is rejected are associated with nonlinearity (i.e., an ESTAR or LSTAR model) as opposed to structural change (i.e., a TVAR model). For the series associated with rejection of linearity, we perform the testing sequence described by Teräsvirta in an attempt to identify whether an ESTAR or LSTAR specification is more appropriate. These results are also reported in Table 2. Accordingly, there are fourteen occasions when an LSTAR model is identified and eight instances where and ESTAR is called for.[5] At this stage of the analysis there seems to be only limited support for the PS hypothesis. Of course it is necessary to attempt to fit appropriate nonlinear models for the eighteen commodity price series for which the linear unit root model is rejected. This is the task to which we now turn.

Estimated STAR–Type Models

The preliminary investigation of STAR or TVAR models for commodity prices begins by attempting to fit the specification called for in Table 2. This is accomplished by using a nonlinear algorithm to estimate each model's parameters, including those that characterize the relevant transition function (van Dijk, Teräsvirta, and Franses, 2003). Of course model estimation is only a preliminary part of the modeling cycle used to fit and assess the performance of the fitted models. Specifically, we employ the diagnostic methods, in the form of LM tests, described by Eitrheim and Teräsvirta (1996) to evaluate the estimated model for: (i) remaining

[4] Among the commodities for which the linear unit root model is not rejected, each is, with the single exception of coffee, associated with a negative drift term. In every instance, however, this term is statistically insignificant at any usual level. We therefore conclude that there is little support for the Prebisch–Singer hypothesis for these commodities.

[5] The total is greater than 18 because, as already noted, for several commodities we consider more than one candidate delay variable.

additive nonlinearity, that is, for specifications similar to that described in (13); and (ii) remaining autocorrelation.[6] In order to conduct tests for remaining nonlinearity, we reserve the first six observations, such that we use $s_t = \Delta y_{t-1}, \ldots, \Delta y_{t-6}$ as candidate transition variables in testing for remaining nonlinearity. The result is there are 92 observations available for model estimation and diagnostic testing. To conserve space, intermediate diagnostic test results are not reported. For each fitted model the estimated speed–of–adjustment parameters, the $\widehat{\gamma}_i$'s, and the estimated location parameters, the $\widehat{c}_i$'s, along with asymptotic standard errors, are reported in Table 3. A summary of diagnostic test results for the final version of the fitted models are recorded in Table 4. Plots of the estimated transition functions, both with respect to the identified transition variable and with respect to time, are displayed in, respectively, the left–hand and right–hand panels of Figure 1.

To begin, initial results revealed that in a handful of instances STAR–type models were inappropriate. Specifically, for cocoa, copper, rubber, and wool the estimated nonlinear models fail to improve on the fit of their linear counterparts as indicated by the AIC.[7] In each instance this seems to be a result of the identified nonlinearity stemming from a relatively small number of outliers. We therefore restrict our attention to the remaining fourteen commodities.

As indicated in Table 2, sevean of these remaining fourteen commodities are associated with a TV–STAR or three–regime STAR model similar to (13). Specifically, aluminum, cotton, lamb, silver, sugar, timber, and tobacco require additional additive components to adequately characterize the data. Of these, four are TV-STAR models (aluminum, cotton, timber, and tobacco) while the remaining three are three-regime STAR models (lamb, silver, and sugar). As implied in Table 2 and Figure 1, and based on the estimated values for γ, five are associated with TAR–type specifications (beef, hides, silver, sugar, and timber). Only in the case of tin does a TVAR model appear to be an entirely adequate specification.

In every instance in which parameter non–constancy is a feature (i.e., aluminum, cotton, timber, tin, and tobacco), the parameter change is smooth. In the case of aluminum a first–order

[6]Because because overall sample sizes are relatively small, we only consider alternatives to first–order STAR–type model specifications that include a second additive term.

[7]Each of these commodities is associated with a negative but otherwise statistically insignificant drift term at usual levels. Indeed, only in the case of wool does the p–value on the drift term approach conventional levels–0.060 for a one–sided test. There is therefore apparently little support for the PS hypothesis for these commodities as well.

logistic function similar to (4) is utilized. Results reveal that the structural break commences in 1927 and is effectively complete by 1947 (Figure 1). For cotton a restricted version of (13) is estimated wherein $c_1 = c_2 = c_3$. Figure 1 reveals that the structural change begins in 1959, plateaus between 1972 and 1979, and is complete by 1990. A similar specification is used for timber (i.e., a restricted third–order TVAR in the second additive component). As indicated in Figure 1, the structural change initiates in 1929, plateaus between 1948 and 1960, resumes again, and finishes by 1978. In the case of tin, the first-order logistic specification in (4) is also called for. The structural change begins in 1940 and is essentially complete by the early 1960s (Figure 1). Finally, and of some particular interest, tobacco is identified with ESTAR–type structural change, that is, with a TVAR in the second additive component specified according to (5). As revealed in Figure 1, the structural change for tobacco begins in the late 1930s, reaches its nadir in 1963, and returns to its original state by the late 1980s. In part, this may reflect the rapid rise in tobacco use in developed countries beginning in the 1930s, a peak that apparently occurred in early 1960s, and a steady decline in use thereafter.

As indicated in Table 4, the estimated STAR–type models apparently fit the data reasonably well. For example, as indicated by the ratio of the standard error of the fitted STAR–type model to its linear counterpart, that is, by, $\widehat{\sigma}_{NL}/\widehat{\sigma}_L$, most estimated STAR–type models provide a substantial improvement relative to the linear ones (Table 4). In some cases, for example, cotton, lamb, silver, and sugar, the improvement in fit is substantial. In seven instances the error distribution of the estimated residuals departs significantly from normality, specifically, in the case of aluminum, beef, lamb, silver, tea, tin, and zinc. With the exception of zinc, this violation is linked to excess kurtosis. As well, and again with the exception of zinc, there is little evidence of remaining residual autocorrelation at four lags (Table 4).

Diagnostic tests of remaining nonlinearity were also performed where where up to six lags, d, of the transition variable $s_t = \Delta y_{t-d}$ are used as candidates. These results are also recorded in Table 4. There is little evidence of remaining nonlinearity in the estimated STAR–type models. While there is some evidence of remaining nonlinearity at the 5% level for aluminum, beef, cotton, and tobacco, attempts to fit additional nonlinear components for these commodities yielded no significant improvements. Only in the case of cotton is the null hypothesis of parameter constancy clearly rejected. But again, attempts to fit a time varying component to cotton resulted in no improvement in fit as measured by AIC. Overall, the diagnostic test results suggest that the estimated models provide a reasonable fit to the data.

Model Simulations

While the foregoing results provide ample evidence of nonlinearity and, in some cases, parameter nonconstancy for a relatively large number of commodities in the sample data, the basic question still remains. Is there evidence that the PS hypothesis holds among the commodities for which STAR–type models are fitted? While there are several ways to investigate this issue, one approach is to examine forward iterations of each model, possibly where stochastic shocks are introduced. That is, what is required are the k–step–ahead forecasts from the estimated models using the ending points of the sample data as initial values.

There are at least two reasons for performing forward simulations of the estimated models. First, and as already mentioned, such simulations will reveal something about the role of the PS hypothesis among commodities for which STAR–type models have been estimated. But of equal importance, forward extrapolations of the model will reveal something about its dynamic properties. Indeed, a necessary condition for stability of an estimated STAR model is that forward iterations of its "skeleton," that is, the forward iterations that do not include stochastic shocks and therefore result in biased forecasts, either converge to a steady–state or a limit cycle path (Tong, 1990). Alternatively, a necessary and sufficient condition for stability is that the forward iterations of the model obtained when shocks are included, that is, the unbiased forecasts, converge to a stable path (Tong, 1990). In this manner useful information may be obtained about each model's dynamic properties.

For the nonlinear models being considered, that is, for the STAR and TV–STAR models, analytical expressions for the forecasts are not available for forecast horizons $k \geq 2$.[8] Therefore, numerical methods must be employed. To see this, let the candidate nonlinear model, rewritten

[8]Because TVAR models do not involve nonlinearity in lagged values of the dependent variables, forecasts from these models may be obtained analytically, and therefore require no special attention.

in levels form, be represented by

$$y_t \;=\; f(y_{t-1},\ldots,y_{t-p-1};\boldsymbol{\theta})+\varepsilon_t,\; t=1,\ldots,T, \tag{16}$$

where $\boldsymbol{\theta}$ is a parameter vector and $\varepsilon_t \sim iid(0,\sigma^2)$. Multiple–step–ahead forecasts are desired from (16) for the period $T+1,\ldots,T+M$, $M\geq 1$. To begin, the one-step-ahead forecast of y_t may be obtained analytically as

$$\widehat{y}_{T+1|T} \;=\; f(y_{T-1},\ldots,y_{T-p-1};\boldsymbol{\theta}),$$

where the usual assumption $E(\varepsilon_{T+1}|\boldsymbol{\Psi}) \;=\; 0$ has been applied, where $\boldsymbol{\Psi}$ denotes the history $y_T, y_{T-1},\ldots$ of observations on y_t. For forecasts at horizons $k\geq 2$, analytical results are no longer available. For example, suppose that we desire a forecast at horizon $k=2$. We then have

$$\widehat{y}_{T+2|T} \;=\; \int_{-\infty}^{\infty} f(\widehat{y}_{T+1|T}+\varepsilon_{T+1}, y_T,\ldots,y_{T-p};\boldsymbol{\theta})d\varepsilon_{T+1}, \tag{17}$$

To solve (17) numerical integration techniques must be employed. If forecasts at horizons $k>2$ are desired, computing the forecast will involve multidimensional numerical integration. At this point several methods may be applied, including Monte Carlo integration and bootstrapping. Here we use the bootstrap method explored originally by Clements and Smith (1997). The requirement is, of course, that the error terms in (16) be independent.

To implement the bootstrap algorithm we simulate N paths for $y_{T+1}, y_{T+2},\ldots,y_{T+k_{max}}$. For present purposes we set $N=1000$ and $k_{max}=200$. We then obtain forecasts for horizons $k\geq 2$ by averaging across all N paths. At horizon $k=2$, for example, we have

$$\widehat{y}_{T+2|T} \;=\; \frac{1}{N}\sum_{i=1}^{N}\widehat{y}^{\,i}_{T+1} \;=\; \frac{1}{N}\sum_{i=1}^{N} f(\widehat{y}_{T+1|T}+\widehat{\varepsilon}_i, y_T,\ldots,y_{T-p};\boldsymbol{\theta}),$$

where $\widehat{\varepsilon}_i$ denotes an estimated residual from (16) sampled with replacement. A naïve forecast or forward simulation may be obtained by simply setting $N=1$ and $\widehat{\varepsilon}_i=0$. The latter amounts to nothing more than a deterministic extrapolation of the so–called “skeleton” of the model.

The forward simulations obtained for each of the fourteen STAR–type models, along with the historical sample data, are presented in Figure 2. In the case of silver, sugar, tea, and lamb, the necessary but not sufficient condition for dynamic stability is satisfied. These results are not at all surprising for silver. As noted by Escribano and Granger (1998), this series is dominated by a bubble during the 1979–80 period. See also Figure 2. From June, 1979 to March, 1980, the Hunt brothers, of Texas, apparently were able to corner the silver market. This bubble period is apparently associated with unstable dynamics. The bootstrap simulations obtained for lamb display a somewhat slow tendency toward explosive behavior; the forecast confidence intervals move well beyond the observed range of the data rather quickly–after about 30 or so forward iterations. As illustrated in Figure 2, the nonlinear dynamics associated with lamb prices are apparently dominated by a steep rise during WWII and a subsequent sharp decline in the early and mid 1950s. In any case, for this group of four commodities only the naïve forecasts are plotted. As well, because tin is associated exclusively with a time–varying autoregression, the forecast results may be obtained analytically–no stochastic simulations are required.

A general conclusion is that for silver, sugar, tea, and lamb, the nonlinear behavior is apparently rather extraordinary in that locally explosive behavior is a prominent feature. In the case of tea, for example, the characteristic polynomial associated with regime $G(.) = 1$ has a dominant real root of 1.40. Even though this regime is observed rather infrequently in the data, see Figure 1, it seems that the model is unable to break free of this dynamic once stochastic shocks are included. Similar results apply for silver, sugar, and lamb. The overall implication is that the estimated STAR models for these commodities might be useful for short–run forecasting, but certainly not for obtaining longer–term predictions. See Hall, Skalin, and Teräsvirta (2001) for a similar example in the context of a nonlinear model of El Niño events.

For the remaining nine commodities, that is, for aluminum, beef, cotton, hides, lead, palmoil, timber, tobacco, and zinc, plots of the bootstrap–based forecasts along with approximate $\pm 2\sigma$ confidence bands are reported in Figure 2.[9] The predictions (both biased and unbiased) plotted in Figure 2 reveal a lack of support in every instance for the PS hypothesis. That is, there is no

[9] As noted by Teräsvirta, van Dijk, and Mederios (2005), a direct benefit of using the bootstrap–based approach to obtain forecasts is that information on the forecast density at each horizon $k \geq 2$ is readily available as a byproduct.

evidence that forward simulations of the estimated STAR–type models results in a continued deterioration of the terms of trade. In part these results are as expected given that the estimated models include lagged level terms. But even so, the TV–STAR models in particular do not require *a priori* that the parameter change be completed by the end of the sample period. And yet this seems to be the case, as depicted in both Figures 1 and 2.

Conclusions

An issue of continuing interest to development and international economists is the prediction, based on the Prebisch–Singer hypothesis, that commodity prices will continue to decline relative to the price of finished or manufactured goods. Numerous studies have in recent years sought to obtain empirical evidence either for or against this basic conjecture. See, for example, Grilli and Yang (1988), Cuddington and Urzua (1989), and Powell (1991). As well, Kim et al. (2003) discuss the need to carefully distinguish between deterministic versus stochastic trends, while others, including Zanias (2005) and Kellard and Wohar (2006), have examined models that allow for one or more trend breaks. Finally, Persson and Teräsvirta (2003) examined the PS hypothesis in the context of a nonlinear time series model. The emerging evidence based on the application of more modern techniques often militates against the PS hypothesis.

Several issues left unexplored in prior research include more formal tests of the linear unit root model against nonlinear or time–varying alternatives, and the specification, estimation, and testing of STAR–type models where called for. Surprisingly, we find that the linear unit root model is rejected in favor of a STAR–type alternative for eighteen of the 24 commodities investigated. Somewhat surprisingly, of these, seventeen were associated with nonlinearity as a viable alternative while only in the case of tin was a TVAR model identified. Of interest is that, with the exception of maize, jute, and rice, all of the commodities associated with a rejection of the linear unit root model were also identified previously by Kellard and Wohar (2006) as being associated with stationary trend break models. Of course Kellard and Wohar (2006) did not consider nonlinear alternatives, as we do here. Among other things it seems that with the relatively small sample sizes that are available it may be difficult to distinguish between trend stationarity (when breaks are incorporated) and nonlinearity.

Of the eighteen commodities for which linearity was rejected, we were able to successfully fit STAR–type models in fourteen instances. Of these, one commodity is associated exclusively with time–varying parameters (tin) while the remaining commodities are best characterized by STAR or additive TV–STAR models. Simulating the models along forward trajectories, either in a stochastic or a deterministic environment, suggest very limited support for the PS hypothesis. At this stage any associated policy implications are not so clear. In part this is because the big peaks and valleys observed in many of the commodity prices seem to be adequately characterized by nonlinearity. That said, it is essentially impossible to predict the the size and direction of any future shocks, and therefore impossible to know with any precision the precise trajectory that relative commodity prices might follow in future. For this reason it will likely remain difficult for developing countries to anticipate how and when to intervene in primary commodity markets, as well as to know which policies to pursue to enhance export earnings.

References

Andrews, D.W.K., Ploberger, W. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62, 1383–1414.

Ardeni, P.G., Wright, B. "The Prebisch–Singer hypothesis: a reappraisal independent of stationarity hypotheses." *The Economic Journal*, 102, 803–812.

Banerjee, A., Lumsdaine, R.L., Stoch, J.H. 1992. "Recursive and sequential tests of the unit–root and trend–break hypotheses: theory and international evidence." *Journal of Business and Economic Statistics*, 10, 271–287.

Caner, M., Hansen, B.E. 2001. "Threshold autoregression with a unit root." *Econometrica*, 69, 1555–1596.

Clements, M.P., Smith J. 1997. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 13, 463–475.

Cuddington, J.T., Urzua, C.M. 1989. "Trends and cycles in the net barter terms of trade: a new approach." *Economic Journal*, 99, 426–442.

Davies, R.B. 1977. "Hypothesis testing when a nuisance parameter is present only under the alternative." *Biometrika*, 64, 247–254.

Davies, R.B. 1987. "Hypothesis testing when a nuisance parameter is present only under the alternative." *Biometrika*, 74, 33–43.

Dickey, D.A., Fuller, W.A. 1979. "Distributions of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association*, 74, 427–431.

Dickey, D.A., Fuller, W.A. 1981. "Likelihood ratio statistics for autoregressive time series with a unit root." *Econometrica*, 49, 1057–1072.

Eitrheim, Ø., Teräsvirta, T. 1996, "Testing the adequacy of smooth transition autoregressive models." *Journal of Econometrics*, 74, 59–75.

Eklund, B. 2003. "Testing the unit root hypothesis against the logistic smooth transition autoregressive model." *Stockholm School of Economics, SSE/EFI Working Paper Series in Economics and Finance*, No. 546.

Enders, W., Granger, C.W.J. 1998. "Unit root tests and asymmetric adjustment with an example using the the term structure of the interest rate." *Journal of Business and Economic Statistics*, 16, 304–311.

Escribano, A., Granger, C.W.J. 1998. "Investigating the relationship between gold and silver prices." *Journal of Forecasting*, 17, 81–107.

Grilli, E.R., Yang, C. 1988. "Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries: what the long run shows." *World Bank Economic Review*, 2, 1–47.

Hall, A. 1994. "Testing for a unit root in time series with pretest data–based model selection." *Journal of Business and Economic Statistics*, 12, 461–470.

Hall, A.D, Skalin, J., Teräsvirta, T. 2001. "A nonlinear time series model of El Niño." *Environmental Modelling & Software*, 16, 139–146.

Hansen, B.E. 1996. "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica*, 64, 413–430.

Holt, M.T., Craig, L.A. 2006. "Nonlinear dynamics and structural change in the U.S. hog–corn cycle: a time–varying STAR approach." *American Journal of Agricultural Economics*, 88, 215–233.

Jarque, C.M., Bera, A.K. 1980. "Efficient tests for normality, homoskedasticity and serial independence of regression residuals." *Economics Letters*, 6, 255–259.

Kapetanios, G., Shin. Y., Snell, A. 2003. "Testing for a unit root in the nonlinear STAR framework." *Journal of Econometrics*, 112, 359–379.

Kellar, N., Wohar, S. 2006. "On the prevalence of trends in commodity prices." *Journal of Development Economics*, 79, 146–167.

Kim, T–H., Pfaffenzeller, N.E., S., Rayner, T., Newbold, P. 2003. "Testing for linear trend with application to relative primary commodity prices." *Journal of Time Series Analysis*, 24, 539–551.

Leon, J., Soto, R. 1997. "Structural breaks and long-run trends in commodity prices." *Journal of International Development*, 3, 44–57.

Li, H., Maddala, G.S. 1996. "Bootstrapping time series models." *Econometric Reviews*, 15, 115–158.

Lin, C–F.J., Tersvirta, T. 1994. "Testing the constancy of structural parameters against continuous structural change." *Journal of Econometrics*, 62, 211–228.

Lomnicki, Z.A. 1961. "Test for the departure from normality in the case of linear stochastic processes." *Metrika*, 4, 37–62.

Lumsdaine, R.L., Papell, D. 1997. "Multiple trend breaks and the unit root hypothesis." *Review of Economics and Statistics*, 79, 212–218.

Lomnicki, Z.A. 1961. "Test for the departure from normality in the case of linear stochastic processes." *Metrika*, 4, 37–62.

Lundbergh, S., Teräsvirta, T., van Dijk, D. 2003. "Time-varying, smooth transition autoregressive models." *Journal of Business and Economic Statistics*, 21, 104–121.

Luukkonen, R., Saikkonen, P., Teräsvirta, T. 1988. "Testing linearity against smooth transition autoregressive models." *Biometrika*, 75, 491–499.

MacKinnon, J. 2002. "Bootstrap inference in econometrics." *Canadian Journal of Economics*, 35, 615–645.

Newbold, P., Vougas, D. 1996. "Drift in the relative price of primary commodities: a case where we care about unit roots." *Applied Economics*, 28, 654–661.

Ocampo, J.A., Parra, M.A. 2004. "The terms of trade for commodities in the twentieth century." *Econ Working Paper Archive*, International Trade Series No. 0402006.

Perron, P. 1989. "The great crash, the oil price shock and the unit root hypothesis." *Econometrica*, 57, 1361–1401.

Perron, P. 1997. "Further evidence on breaking trend functions in macroeconomic variables." *Journal of Econometrics*, 80, 355–385.

Persson, A., Teräsvirta, T. 2003. "Net barter terms of trade: a smooth transition approach." *Journal of Financial Economics*, 8, 81–97.

Powell, A. 1991. "Commodity and developing country terms of trade: what does the long run show?" *Economic Journal*, 101, 1485–1496.

Prebisch, R. 1950. *The Economic Development of Latin America and its Principal Problems.* United Nations Publications: New York.

Said, S.E., Dickey, D.A. 1984. "Testing for unit roots in autoregressive–moving average models of unknown order." *Biometrika*, 71, 599–607.

Singer, H. 1950. "The distributions of gains between investing and borrowing countries." *American Economic Review*, 40, 473–485.

Teräsvirta, T. 1994. "Specification, estimation, and evaluation of smooth transition autoregressive models." *Journal of the American Statistical Association*, 89, 208–217.

Teräsvirta, T., Anderson, H.M. 1992. "Characterizing nonlinearities in business cycles using smooth transition autoregressive models." *Journal of Applied Econometrics*, 7, S119–S136.

Teräsvirta, T., van Dijk, D., Medeiros, M.C. 2005. "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination." *International Journal of Forecasting*, 21, 755–744.

Tong, H. 1990. *Nonlinear Time Series: A Dynamical Systems Approach.* Oxford University Press: Oxford.

van Dijk, D., Franses, P.H. 1999. "Modeling multiple regimes in the business cycle." *Macroeconomic Dynamics*, 3, 311–340.

van Dijk, D., Teräsvirta, T. , Franses, P.H. 2002. "Smooth transition autoregressive models–a survey of recent developments." *Econometric Reviews*, 21, 1–47.

Zanias, G.P. 2005. "Testing for trends in the terms of trade between primary commodities and manufactured goods." *Journal of Development Economics*, 78, 49–59.

Zivot, E., Andrews, D.W.K. 1992. "Further evidence on the great crash, the oil price shock, and the unit root hypothesis." *Journal of Business and Economic Statistics*, 10, 251–270.

Table 1: Results of Dickey–Fuller Tests Applied to 24 Commodities.

	No Trend under the Alternative			Trend under the Alternative		
Commodity	$\widehat{\rho}$	$t_{\widehat{\mu}}$	p–value	$\widehat{\rho}$	$t_{\widehat{\tau}}$	p–value
Aluminum	0.920	-2.344	0.121	0.864	-2.620	0.244
Bananas	0.919	-1.710	0.414	0.887	-2.248	0.472
Beef	0.923	-2.006	0.290	0.848	-2.547	0.299
Cocoa	0.870	-2.290	0.167	0.865	-2.355	0.395
Coffee	0.838	-2.470	0.107	0.841	-2.372	0.410
Copper	0.851	-2.178	0.164	0.851	-2.149	0.438
Cotton	0.984	-0.365	0.902	0.870	-1.991	0.570
Hides	0.851	-2.169	0.225	0.618	-3.519	0.045**
Jute	0.848	-1.757	0.421	0.799	-2.187	0.503
Lamb	0.918	-1.998	0.274	0.793	-3.025	0.152
Lead	0.847	-1.921	0.296	0.843	-1.636	0.695
Maize	0.955	-0.692	0.866	0.665	-3.017	0.120
Palm Oil	0.926	-1.214	0.688	0.747	-2.691	0.224
Rice	0.942	-1.167	0.665	0.790	-2.598	0.249
Rubber	0.923	-2.097	0.181	0.762	-3.628	0.040**
Silver	0.912	-1.831	0.350	0.898	-2.018	0.564
Sugar	0.834	-1.839	0.372	0.689	-2.646	0.295
Tea	0.928	-1.463	0.524	0.893	-1.990	0.583
Timber	0.902	-2.222	0.174	0.734	-3.507	0.043**
Tin	0.883	-2.086	0.261	0.832	-2.336	0.402
Tobacco	0.913	-2.765	0.032**	0.918	-1.610	0.653
Wheat	0.975	-0.429	0.895	0.635	-3.088	0.072*
Wool	1.020	0.515	0.988	0.887	-1.486	0.609
Zinc	0.560	-3.016	0.026**	0.557	-3.005	0.102

Note: $\widehat{\rho}$ is the estimated root. The test statistics $t_{\widehat{\mu}}$ and $t_{\widehat{\tau}}$ are t–ratios for $(\widehat{\rho} - 1)$, and correspond, respectively, to: (1) the case where the estimated model does not include a trend, and (2); the case where the estimated model does include a linear trend. Columns headed p–value record approximate p–value's based on $B = 999$ bootstrap simulations. A superscripted * indicates significance at the 10% level, ** significance at the 5% level and, *** significance at the 1% level.

Table 2: Results of Testing a Linear Unit Root Model Against STAR or TVAR Alternatives and of Applying Teräsvirta's (1994) Model Selection Sequence.

Commodity	p	$\min[p_{F_{lur}}]$	d	p_{F_4}	p_{F_3}	p_{F_2}	STAR/TVAR
Aluminum	2	0.023**	2	0.185	0.466	0.026	LSTAR
	2	0.028**	6	0.019	0.148	0.160	LSTAR
	2	0.032**	1	0.315	0.009	0.276	ESTAR
Banana	2	0.164					
Beef	1	0.015**	3	0.005	0.022	0.502	LSTAR
Cocoa	2	0.025**	3	0.048	0.184	0.125	LSTAR
Coffee	2	0.118					
Copper	5	0.019**	6	0.005	0.775	0.313	LSTAR
Cotton	3	0.002***	3	0.001	0.050	0.692	LSTAR
	3	0.051*	2	0.618	0.046	0.145	ESTAR
Hides	2	0.028**	2	0.273	0.095	0.035	LSTAR
Jute	4	0.115					
Lamb	5	0.047**	1	0.202	0.019	0.478	ESTAR
Lead	4	0.014**	6	0.490	0.018	0.026	ESTAR
Maize	4	0.312					
Palmoil	5	0.014**	2	0.001	0.194	0.196	LSTAR
Rice	4	0.115					
Rubber	1	0.020**	4	0.003	0.144	0.375	LSTAR
Silver	2	0.001***	2	0.001	0.194	0.196	LSTAR
	2	0.001***	6	0.018	0.394	0.001	LSTAR
Sugar	5	0.022**	1	0.168	0.511	0.061	LSTAR
	5	0.085*	2	0.147	0.985	0.045	LSTAR
Tea	2	0.091*	4	0.950	0.020	0.129	ESTAR
Timber	2	0.009***	1	0.114	0.018	0.061	ESTAR
Tin	3	0.045**	t	0.130	0.319	0.262	TVAR
Tobacco	4	0.004***	4	0.176	0.001	0.231	ESTAR
Wheat	6	0.302					
Wool	4	0.057*	1	0.427	0.064	0.298	ESTAR
Zinc	6	0.093*	5	0.696	0.490	0.038	LSTAR

Note: The column headed p denotes the optimal number of lags in the linear AR model. The column headed $\min[p_{F_{lur}}]$ denotes the minimum p–value of the linear unit root test over delays $d = 1, \ldots, 6$. The column headed d denotes the delay corresponding to $\min[p_{F_{lur}}]$, and columns headed, respectively, p_{F_4}, p_{F_3}, and p_{F_2} correspond to p–values of tests in the model selection sequence. All p–values are obtained by performing $B = 999$ recursive bootstraps of the model's residuals under the respective null hypothesis. The final column indicates whether an LSTAR, ESTAR, or TVAR model is chosen. A superscripted * indicates significance at the 10% level, ** significance at the 5% level, and *** significance at the 1% level.

Table 3: Estimated Speed of Adjustment Parameters, γ, and Location Parameters, c, in the Transition Functions of the Estimated Models.

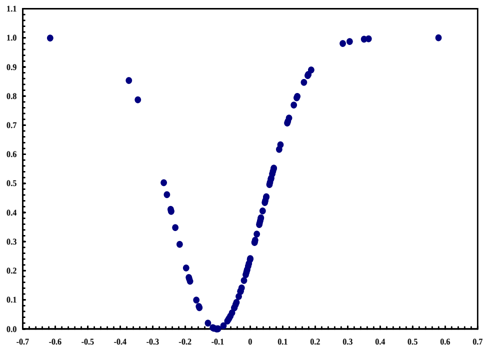
	First Transition Function				Second Transition Function			
Commodity	Type	d	$\widehat{\gamma}_1$	$\widehat{c}_1$	Type	d	$\widehat{\gamma}_2$	$\widehat{c}_2$
Aluminum	ESTAR	1	0.716	-0.102	LTVAR	t	14.531	0.347
			(0.461)	(0.034)			(11.607)	(0.030)
Beef	LSTAR	3	500	0.159				
			–	(0.002)				
Cotton	ESTAR	2	4.341	-0.018	LTVAR	t^3	27.414	0.744
			(2.752)	(0.016)			(11.685)	(0.020)
Hides	LSTAR	2	500	-0.097				
			–	(0.015)				
Lamb	ESTAR	1	0.382	-0.208	LSTAR	2	4.247	0.108
			(0.160)	(0.049)			(2.159)	(0.038)
Lead	ESTAR	6	1.109	-0.052				
			(0.501)	(0.022)				
Palmoil	LSTAR	2	25.945	0.175				
			(19.069)	(0.010)				
Silver	LSTAR	2	264.574	0.154	LSTAR	4	500	0.026
			(613.244)	(0.002)			–	(0.003)
Sugar	LSTAR	1	500	7.20E-05	LSTAR	2	104.708	0.073
			–	(0.002)			(151.382)	(0.004)
Tea	ESTAR	4	0.929	-0.094				
			(0.614)	(0.028)				
Timber	ESTAR	1	0.807	0.062	LTVAR	t^3	7.054	0.521
			(0.557)	(0.019)			(3.454)	(0.038)
Tin	LTVAR	t	11.756	0.510				
			(13.357)	(0.049)				
Tobacco	ESTAR	4	0.756	0.036	ETVAR	t	4.170	0.620
			(0.325)	(0.017)			(2.781)	(0.043)
Zinc	LSTAR	6	500	-0.007				
			–	(0.003)				

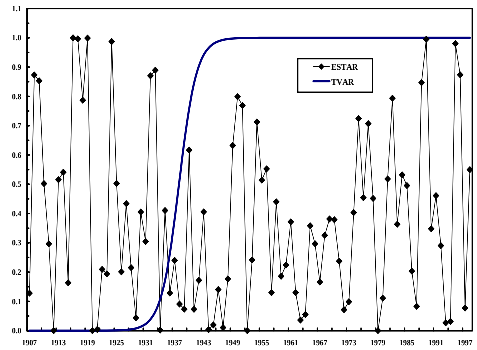
Note: Values in parentheses are asymptotic standard errors. LTVAR denotes a logistic transition function with time as an argument. Alternatively, ETVAR denotes an exponential transition function with time as an argument. Under the column headed d the entry t^3 denotes a restricted third–order LTVAR transition function.

Table 4: Diagnostic Test Results for Estimated STAR–type Models.

									Tests for Remaining Nonlinearity/Parameter Constancy						
Commodity	$\widehat{\sigma}_\varepsilon$	AIC	R^2	$\widehat{\sigma}_{NL}/\widehat{\sigma}_L$	SK	EK	LJB	LM_{AR}	Δy_{t-1}	Δy_{t-2}	Δy_{t-3}	Δy_{t-4}	Δy_{t-5}	Δy_{t-6}	t
Aluminum	0.125	-3.818	0.432	0.841	0.776	3.637	62.59(2.56E-14)	0.072	0.189	0.102	0.388	0.950	0.399	0.032	0.269
Beef	0.209	-2.956	0.180	0.926	0.044	2.043	16.68(2.39E-4)	0.197	0.016	0.466	0.008	0.760	0.277	0.843	0.145
Cotton	0.115	-3.917	0.502	0.777	0.062	-0.211	1.12(0.573)	0.099	0.915	0.058	0.050	0.595	0.494	0.757	0.018
Hides	0.211	-2.891	0.282	0.925	0.009	-0.170	0.25(0.881)	0.764	0.822	0.361	0.472	0.807	0.776	0.887	0.540
Lamb	0.173	-2.964	0.501	0.781	0.201	1.901	16.94(2.10E-4)	0.207	0.495	0.176	0.365	0.152	0.133	0.344	0.889
Lead	0.140	-3.634	0.353	0.850	0.001	-0.126	0.079(0.962)	0.486	0.610	0.638	0.186	0.617	0.692	0.174	0.095
Palmoil	0.178	-3.109	0.380	0.875	0.003	0.793	2.46(0.293)	0.698	0.104	0.044	0.363	0.515	0.610	0.061	0.073
Silver	0.137	-3.630	0.474	0.774	0.241	1.300	10.17(6.20E-3)	0.192	0.375	0.515	0.786	0.987	0.302	0.831	0.999
Sugar	0.214	-2.538	0.630	0.683	0.175	0.633	4.22(0.121)	0.431	0.817	0.069	0.358	0.735	0.700	0.111	0.184
Tea	0.140	-3.720	0.261	0.902	0.009	1.614	10.13(6.33E-8)	0.159	0.082	0.077	0.944	0.315	0.996	0.270	0.764
Timber	0.107	-4.120	0.383	0.821	0.001	0.826	2.63(0.269)	0.478	0.675	0.968	0.469	0.417	0.846	0.714	0.364
Tin	0.149	-3.545	0.260	0.915	0.347	2.681	32.87(7.27E-8)	0.671	0.351	0.869	0.671	0.124	0.595	0.502	0.163
Tobacco	0.077	-4.653	0.489	0.830	0.041	0.494	1.56(0.457)	0.573	0.045	0.499	0.023	0.453	0.048	0.084	0.086
Zinc	0.146	-3.463	0.449	0.847	2.114	6.424	190.6(4.06E-42)	0.026	0.596	0.640	0.028	0.259	0.395	0.509	0.288

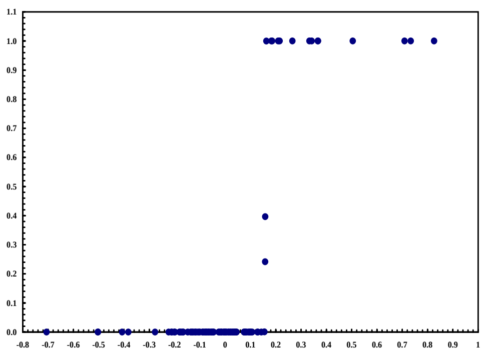
Note: $\widehat{\sigma}_\varepsilon$ is the residual standard error; AIC is Akaike information criterion; R^2 is the unadjusted R^2, and $\widehat{\sigma}_{NL}/\widehat{\sigma}_L$ is the ratio of the residual standard error from the respective STAR–type model relative to the linear AR model. As well, SK is skewness, EK is excess kurtosis, and LJB is the Lomnicki– Jarque–Bera testof normality of residuals, with asymptotic p–values in parentheses. LM_{AR} denotes the F variant of Eitrheim and Teräsvirta's (1996) LM test of no remaining autocorrelation in the residuals based on four lags. Remaining columns report p–values for F variants of Eitrheim and Teräsvirta's (1996) LM tests for remaining nonlinearity, $(\Delta y_{t-1}, \ldots, \Delta y_{t-6})$, and parameter constancy, t.

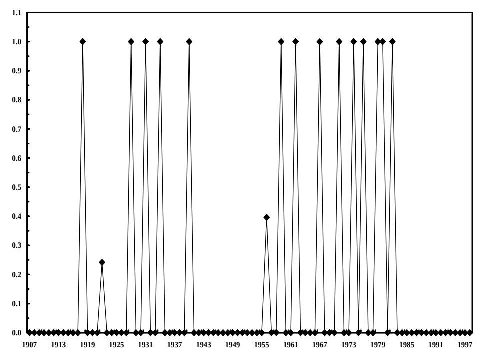




Aluminum

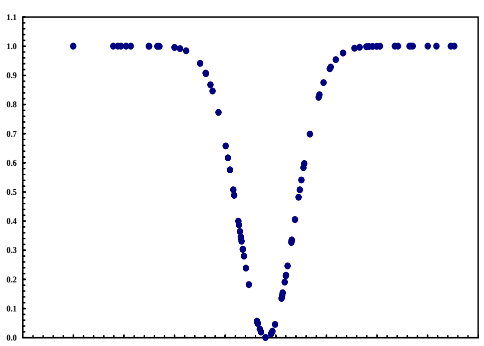
Aluminum

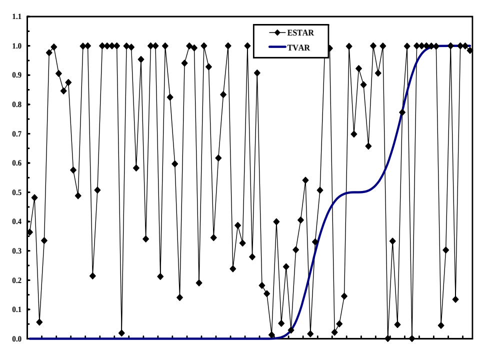




Beef

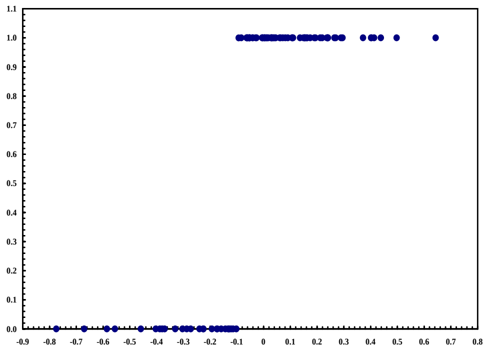
Beef



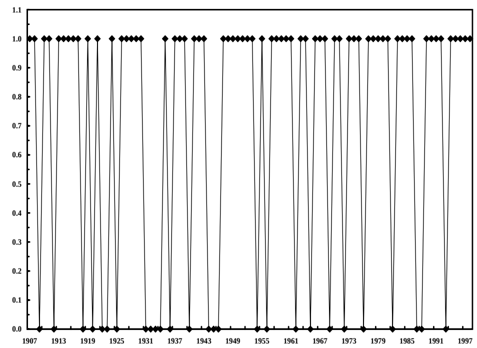


Cotton

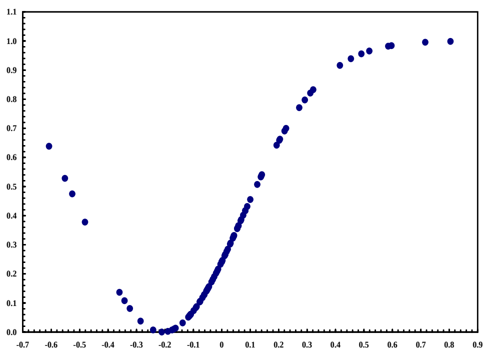
Cotton



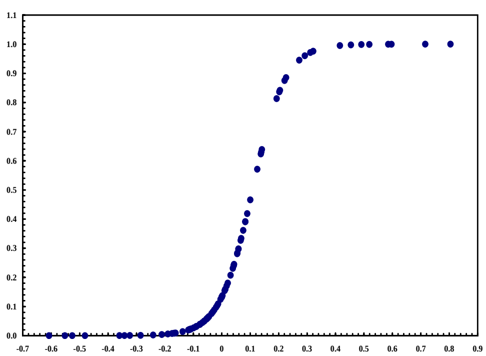
Hides



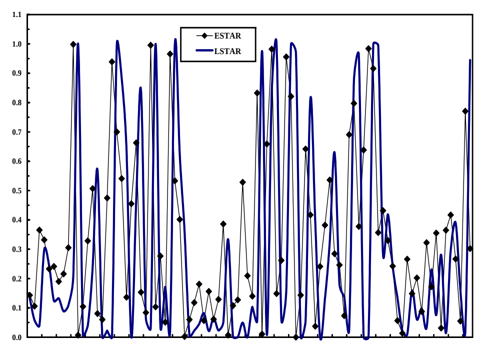
Hides



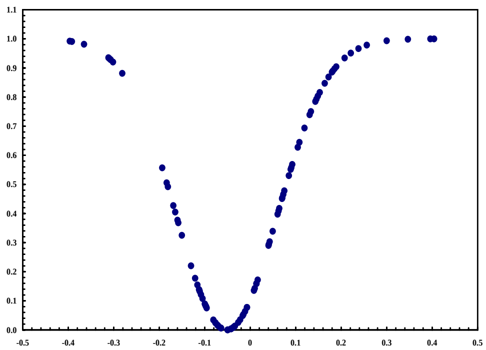
Lamb



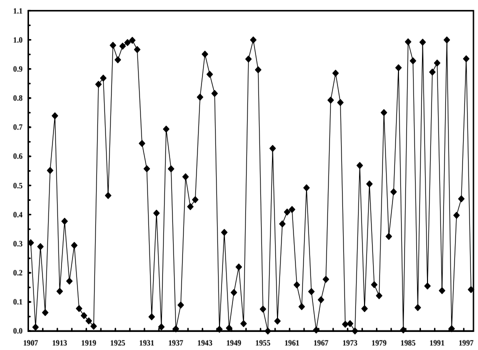
Lamb



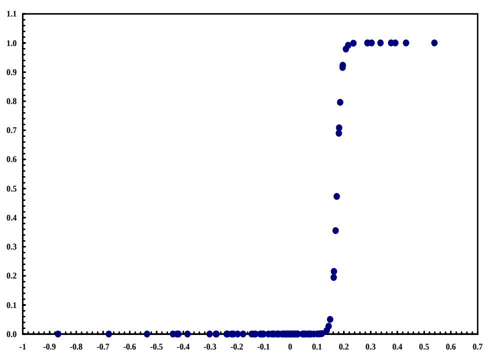
Lamb



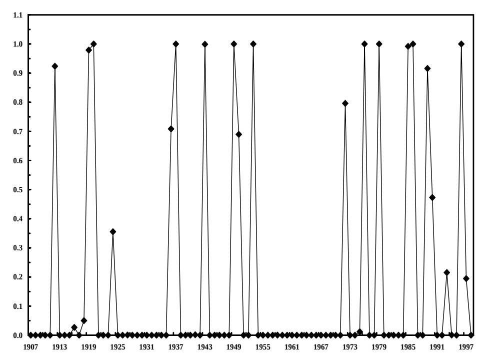
Lead



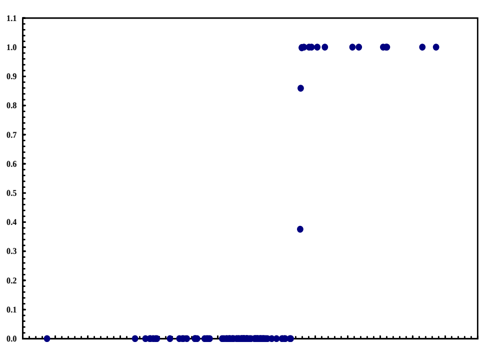
Lead



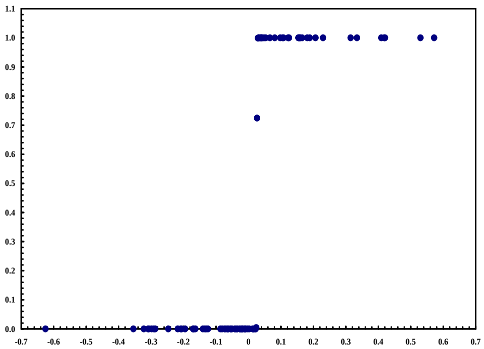
Palmoil



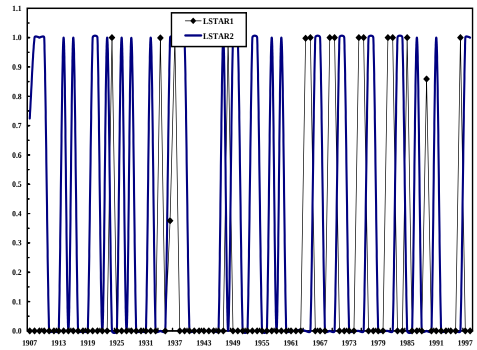
Palmoil



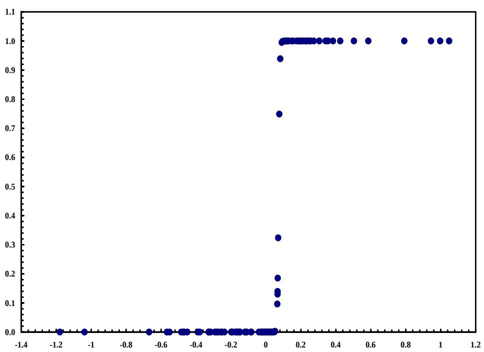
Silver



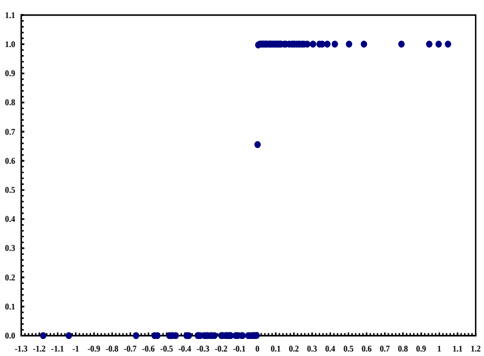
Silver



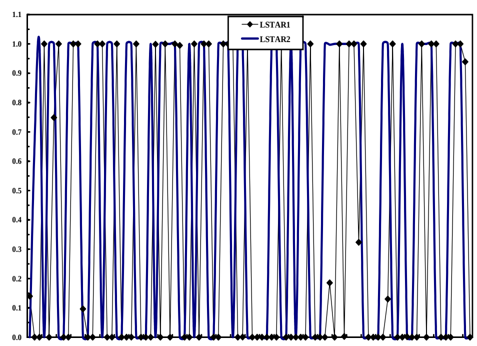
Silver



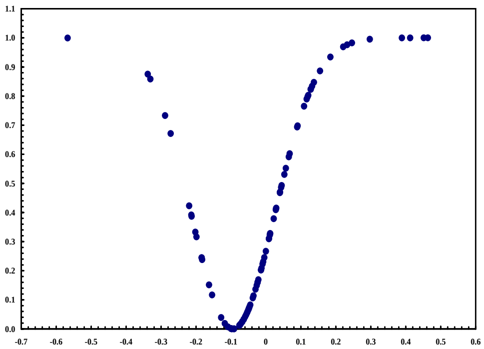
Sugar



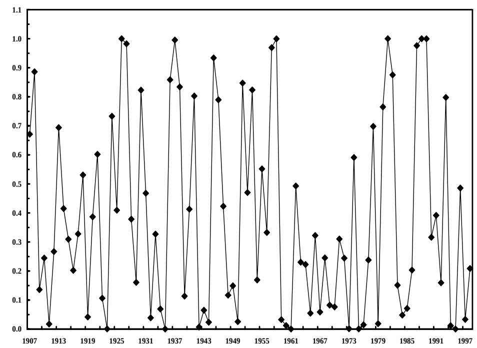
Sugar



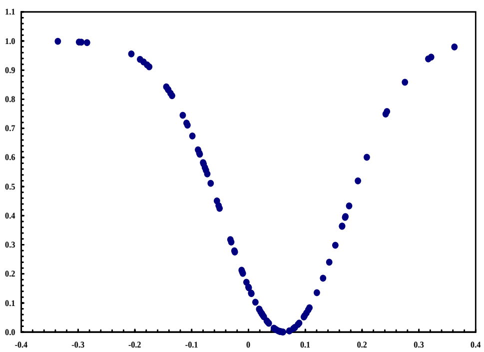
Aluminum



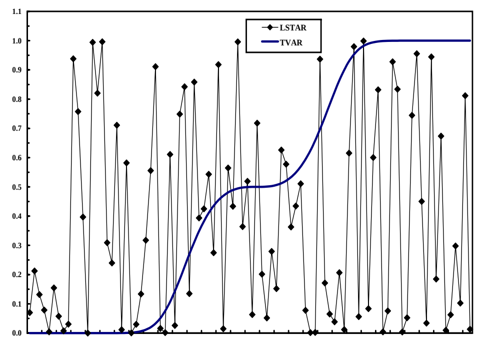
Tea



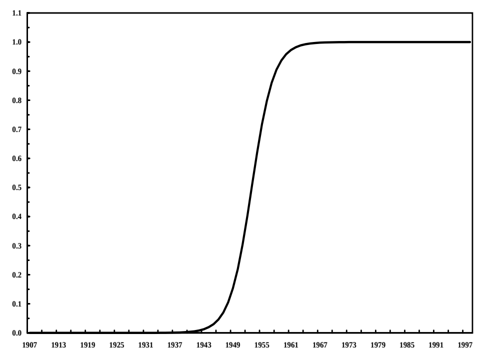
Tea



Timber



Timber



Tin

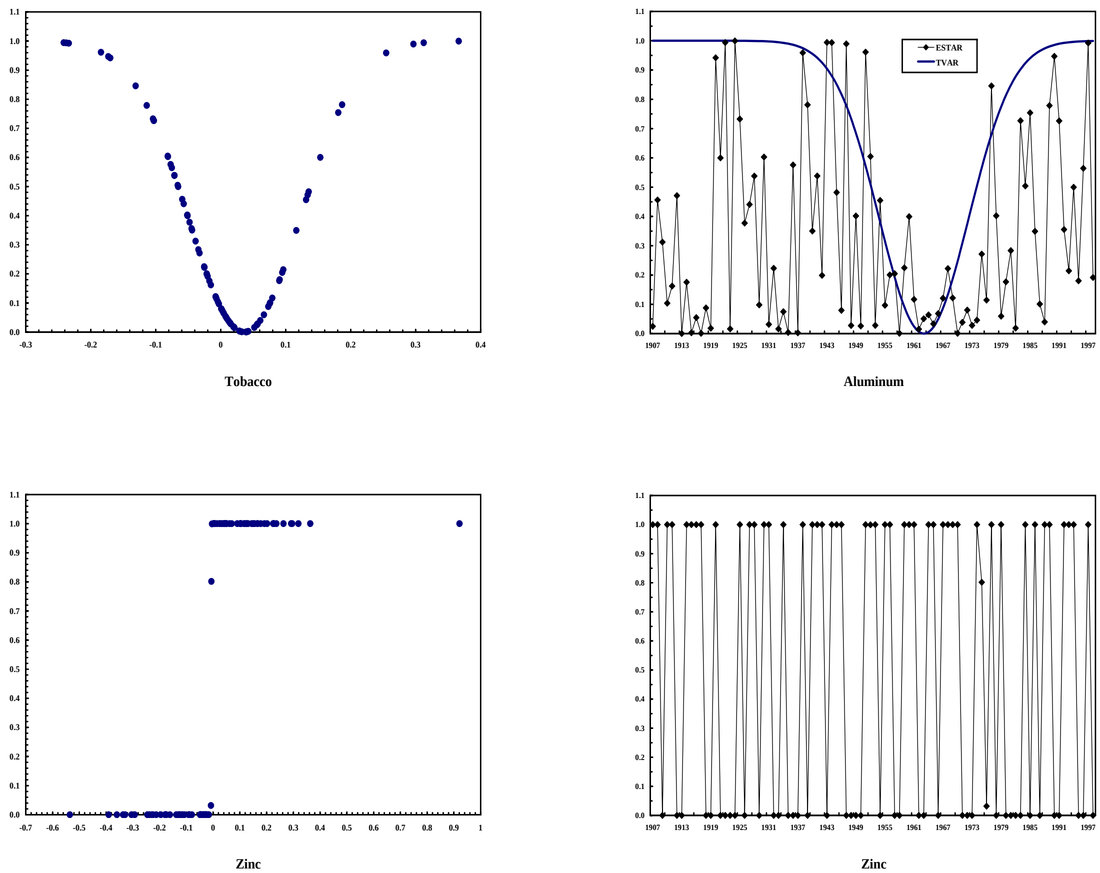
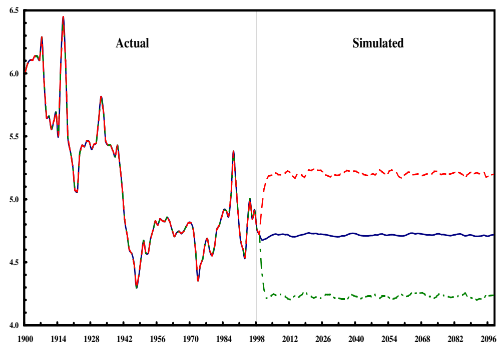
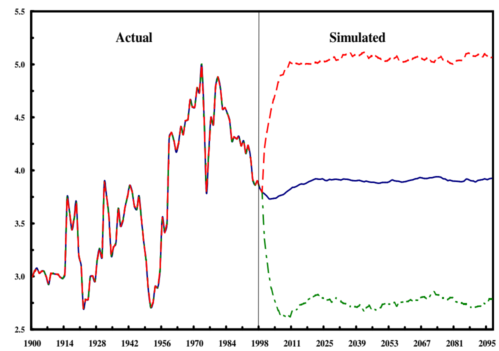


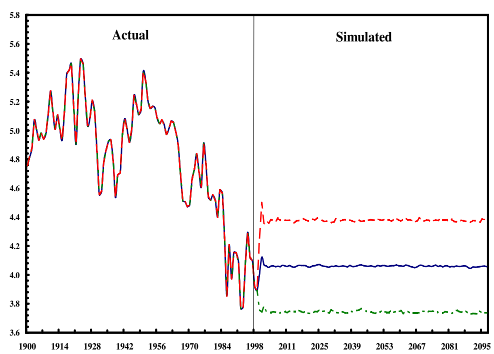
Figure 1: Transition Functions versus the Respective Transition Variable (left–hand column) and Transition Functions Over Time (right–hand column) for Fourteen Commodities.



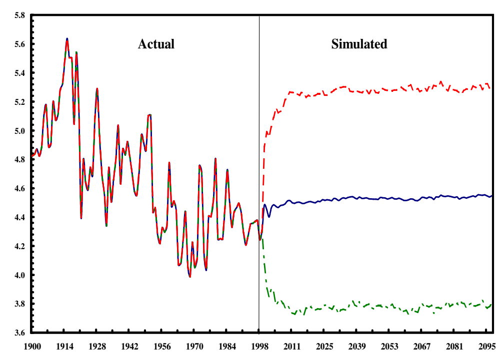
Aluminum



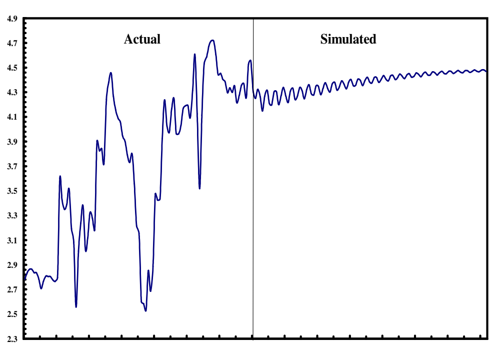
Beef



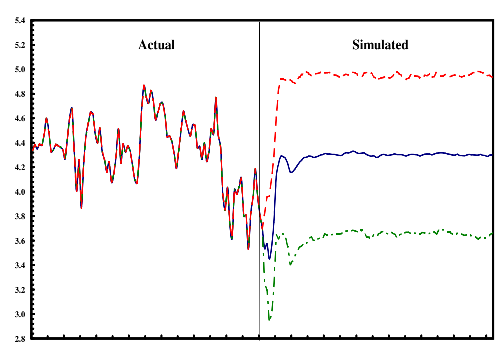
Cotton



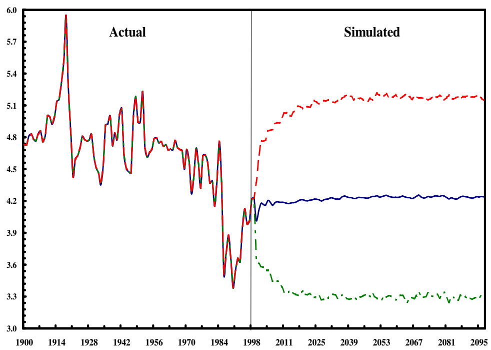
Hides



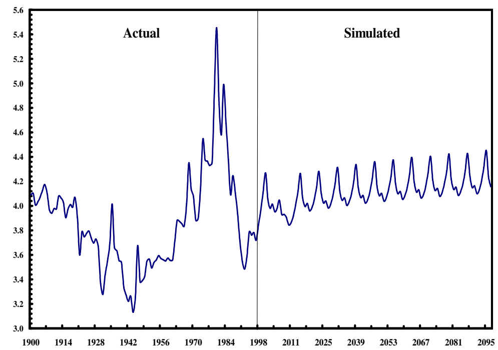
Lamb



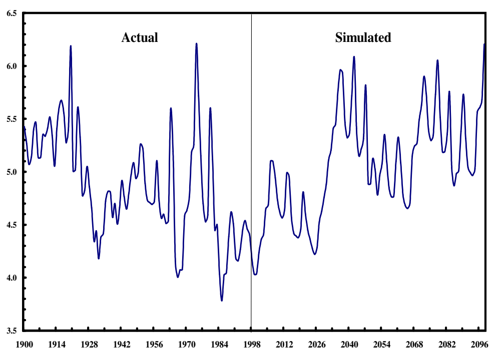
Lead



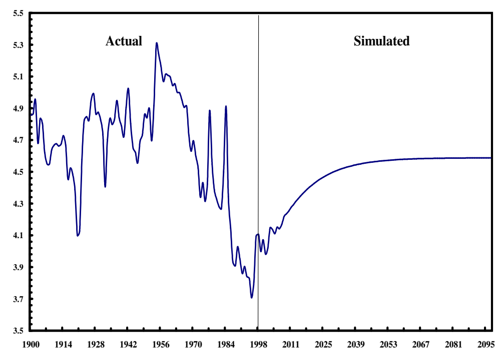
Palmoil



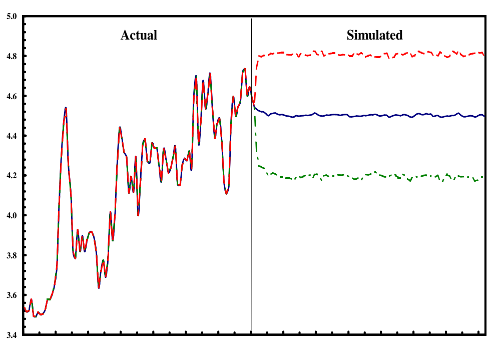
Silver



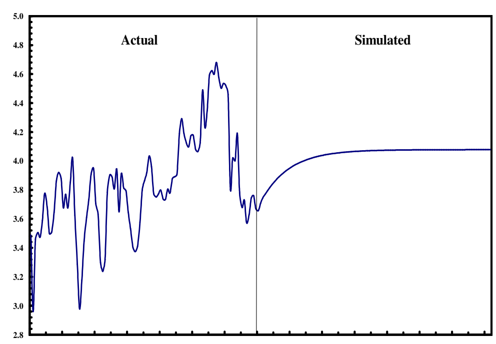
Sugar



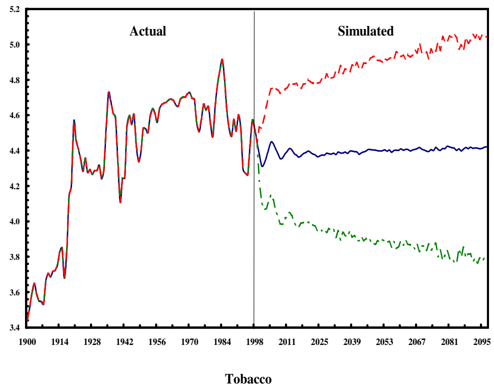
Tea



Timber



Tin



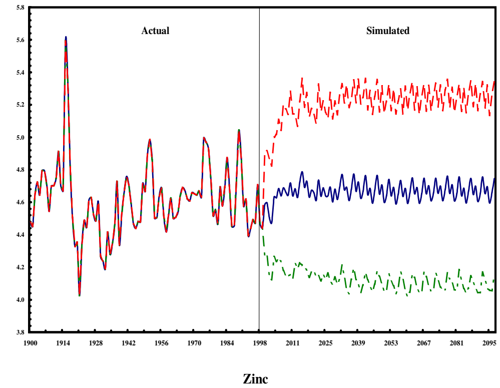


Figure 2: Actual (1900–1998) and Simulated (1999–2097) Commodity Price Data for Fourteen Commodities.