Eurasian Journal of Economics and Finance, 9(2), 2021, 107-127
DOI: 10.15604/ejef.2021.09.02.004

EURASIAN JOURNAL OF ECONOMICS AND FINANCE

www.eurasianpublications.com

IN- AND OUT-OF-SAMPLE PERFORMANCE OF NONLINEAR MODELS IN INTERNATIONAL PRICE DIFFERENTIAL FORECASTING IN A COMMODITY COUNTRY FRAMEWORK †

Nicola Rubino
University of Huelva, Spain
Email: nicola.rubino.2@gmail.com

Received: February 23, 2021 Accepted: May 24, 2021

Abstract

This paper presents an analysis of a group of small commodity-exporting countries' price differentials relative to the US dollar. Using unrestricted self-exciting threshold autoregressive models (SETAR), we evaluate the sixteen national Consumer Price Indexes (CPI) differentials relative to the US dollar CPI. Out-of-sample forecast accuracy is estimated through calculation of mean absolute errors measures based on the monthly rolling window and recursive forecasts, and this estimation is extended to three additional models, namely a logistic smooth transition regression (LSTAR), an additive nonlinear autoregressive model (AAR), and a simple neural network model (NNET). Our preliminary results confirm the presence of some form of nonlinearity in most of the analyzed countries. The parsimonious AR(1) model does not appear to perform any worse than any nonlinear model in the rolling sample exercise. However, in terms of a long-run equilibrium driven by purchasing power parity, its validity is undermined by the results of the recursive estimates and the outcome of the Diebold-Mariano type tests, which favor generally the Heckscher commodity points theory. As a policy advice to commodity-exporting countries, we find no apparent reason to suggest commodity export price pegging as a generalized foreign exchange policy.

Keywords: Transition Regression Model, Real Exchange Rate, Nonlinearities, Price Differentials, Purchasing Power Parity (PPP), Commodity Points

JEL Classifications: C22, C32, F31

1. Introduction

This paper analyzes the short-run adjustment properties of a series of international price differentials in a set of emerging commodity-exporting countries under the light of three competing theories, which embody the last twenty-one years of applied analysis on the convergence of price differentials and exchange rates to a univariate equilibrium in observed components model analyses: the classical (linear) purchasing power parity (PPP) theory; the commodity points theory, that is, the Heckscher commodity points theory seen as the studies related to the transaction cost-based interpretation of nonlinearities of the real exchange adjustment path to its equilibrium; the smooth commodity points theory, that is, the aforementioned theory, only seen

† This paper is based on the Chapter 2 of the Phd thesis entitled "Three Essays on Commodity Prices", authored by Nicola Rubino and defended at the University of Barcelona, Faculty of Economics.

on the time-delaying retrospective that agents might somewhat take their time to adjust their expectations to changes in the exchange rate behavior.[1] 1 As we test for the above theories, we unavoidably also tested for the puzzle of exchange rate slow speed of mean reversion and consequent failure to produce reliable forecasts.[2]

Our paper makes two contributions: on the one hand, our work focuses on price differentials behavior in a wide set of emerging countries[3], which have been largely disregarded up until now due to data limitations, on the other hand, it analyzes the performance of a benchmark linear model [an AR(1)] compared to an exhaustive set of nonlinear models to reach some decisive conclusions on the possibility of forecasting the price differentials in commodity-exporting countries, followed by a final policymaking suggestion.

Our work features a univariate analysis of the price differentials of an ad-hoc selected group of export-specialized countries with respect to the US dollar. Employing a variety of time series models (standard univariate linear regressions, smooth and discrete transition autoregressive models, simple neural network models, and additive nonlinear auto-regressive models), we examine and compare the behavior of sixteen currencies with respect to the US dollar and evaluate the models through standard forecast error metrics such as the mean absolute percentage error (MAPE) and the in-sample Bayesian information criterion (BIC) to look for a dominant class of model and consequentially the best-suited theory describing the national currency/US dollar differential.

Our main argument is that, contrary to the spirit of the past empirical research on the short-run adjustment of price differentials in industrialized countries, in- and out-of-sample performance of nonlinear models, as implied by the Heckscher commodity points theory, might outperform the standard convergence to equilibrium implied by the usual purchasing power parity benchmarks in countries specialized in commodity exports, thus offering a specific solution to the Penn effect in such countries. Secondly, we argue that the above hypothesis might help policy in forecasting price differentials to set more efficient policies for the mentioned commodity-exporting countries. To tackle such arguments, our paper presents itself as a fundamental in-sample and out-of-sample forecasting exercise.

Our findings suggest that the parsimonious AR(1) model used as a benchmark for our exercises does not appear to perform any worse than any nonlinear model in the rolling sample exercise. However, a set of Diebold-Mariano type tests, which we employed, would generally favor the Heckscher commodity points theory. Given the nature of our results, and as a policy advice to small commodity exporters, we find no apparent reason to suggest commodity export price pegging as a generalized foreign exchange policy.

The paper is, thus, organized as follows: In Section 2, we present a brief literature review related to our analysis. In Section 3, we underline the theoretical framework. In Section 3.1, we present the data of the paper and how it was treated. In Section 4, we present the nonlinear methods we employed for our in-sample estimates and out-of-sample exercise. In Section 4.1, we illustrate the stationarity tests we used as well as their results. In Section 4.2, we discuss the results of the F-tests on the feasibility of a nonlinear representation of the international price differentials. In Section 5, we present the estimated values of the attractor/error correction coefficients, calculate their half-lives, and finally compare the in-sample performance of models with an information criterion and a measure of forecast error. In Section 5.1, we run a rolling window forecast over twenty periods forecasting horizon to compare the out-of-sample performance of the models to the benchmark AR(1) specification. Finally, Section 6 concludes the paper.

[1] See Rogoff (1996) for a complete overview.

[2] Some relevant theoretical studies suggest that a smooth adjustment may be more apt at capturing transaction costs when they are proportional (as in Taylor *et al.* 2001)).

[3] The appendix reports a graphical rendition of the time-series for Kenya, together with the detrended series and the out-of-sample graphic output of the forecasts (Figure 1 to 3). For an overview of the descriptive statistics of the series, the reader is invited to check Table 17 in the Appendix.

2. Literature review
2.1. Nonlinearities in exchange rates

The literature, based on the nonlinear modeling of the exchange rates, has been spanning the last twenty years of academic research. Among the many seminal contributions to that branch of international economics seeking an answer to the exchange rate slow return to its mean despite its very high (and basically intra-day) volatility, we recall the seminal work of Balke and Fomby (1997) on threshold cointegration, the seminal contribution of Obstfeld and Taylor (1997) on equilibrium and band threshold autoregression (equilibrium or band TAR) applications to the US CPI-based price differential mean reversion in a group of six advanced countries, and lastly the important extension to the former study given by the smooth transition autoregressive application by Michael *et al.* (1997). Although a lot has been said on price differential adjustment and exchange rate behavior about its equilibrium,[4] not much attention has been given to possible alternative nonlinear solutions to price differential forecasting in small commodity-exporting countries except for a few major linear behavioral equilibrium applications[5]. As for the most recent literature, we should perhaps cite the panel data linear applications of Bodart *et al.* (2012) and Bodart *et al.* (2015), who updated the results from Cashin *et al.* (2004) with an analysis of discrete threshold levels for exported commodity prices, Coudert *et al.* (2015), who explored the nexus between real exchange rates and prices in a nonlinear panel setting, and Ricci *et al.* (2013), who employed once more a linear panel setting to check for the elasticities of the exchange rate with respect to prices in an exhaustive panel of 48 industrial and emerging countries. Our paper bridges the gap between the two research fields of international economics as it attempts to shed some light on possible nonlinearities in the speed of mean reversion of national to US CPI differentials in a set of countries specialized in exporting a single commodity with an export share of at least twenty percent over the total volume of exports.[6]

2.1.2. Recent advances in forecasting

The random walk (with drift) has always been considered the best available predictive model in terms of out-of-sample forecasting for nominal and real exchange rates (Meese and Rogoff, 1983). The seminar result of the former authors is perhaps contradictory to purchasing power parity, which would have to rely on a long-time span rather than high frequencies and many observations (see Harvey *et al.* (2010), who studied the secular decline in commodity prices for the past three centuries). Indeed, modern forecasting is nowadays pretty much centered on the exact choice of the structure of covariates (i.e., net foreign assets and Taylor rules or a univariate approach), the frequency of the analysis, and the best performing forecasting mechanism (i.e., rolling versus recursive) (Rossi, 2013). Recently, researchers have underlined that exchange rates series could be used to predict commodity at quarterly frequencies (Chen *et al.* 2010), while energy prices (oil) would be able to predict real exchange rates at even lower ones (Ferraro *et al.* 2015). Perhaps the main common factor which unites the mentioned authors and the main bulk of the literature on forecasting exchange rates and price differentials stands on the ground of linearity. Our paper contributes to this branch of the forecasting literature by comparing alternative nonlinear methodologies to an autoregressive benchmark in a set of commodity countries that would be most probably present nonlinearities in their price differential behavior.

[4] One very recent example being Allen *et al.* (2016), which focused on out-of-sample forecasting performance in a series of nonlinear univariate specifications for the exchange rate for a set of six hard currencies.
[5] We cite Cashin *et al.* (2004), who focused on a large group of commodity exporters, and Chen and Rogoff (2003), who considered commodity exported but limited their attention to three big, developed economies.
[6] Note that not a lot of literature has covered emerging/exporting countries yet (see again the seminal paper by Cashin *et al.* (2004)). Furthermore, past literature would focus on convergence, not forecasting.

2.1.3. Policy implications

The possibility of forecasting price differentials more efficiently makes the real exchange rate an objective variable for policymakers. Confirmation of the Heckscher commodity points theory could make the difference between a flexible policy stance or the idea of pegging the nominal exchange rate to the most exported leading commodity price (or index) to eliminate risks deriving from unwanted fluctuations. This is especially true for small exporting countries focusing on the export of a non-diversified set of goods or a single commodity. In a sense, the adoption of a flexible exchange rate regime based on the exported commodity price would reap the advantages of anchoring the exchange rate to a nominal anchor and at the same time allow for the degree of insulation from terms of trade shocks that a standard flexible exchange rate should guarantee. Among the proponents of such ideas, it is worth mentioning the seminal papers of Frankel (2005) and Frankel and Saiki (2002).

3. Theoretical framework

Our log-differential prices represent an exchange rate, defined as the cost of a basket of goods relative to the same basket between two countries once such basket has been converted to a common numeraire. As far as the definition we employ goes, our numeraire is not a good (thus implying a real exchange rate) but a common currency. The differentials we analyze are obtained by deflating nominal indexes by the US Consumer Price Index expressed in dollars (Equation 1).

$$\frac{CPI_i}{CPI_{US}} \tag{1}$$

Introducing the purchasing power parity theory and using for simplicity home and away price notation, P_1 and P_2^x, the relationship between the two price baskets in levels and expressed in real terms will thus be (Equation 2).

$$P_1 = \varepsilon P_2^* \tag{2}$$

where; ε represents the nominal bilateral exchange rate. The absolute version of the purchasing power parity assumes that exchange rates should be equal to one or tend to return very quickly to such equilibrium in the long run. As such, for this absolute version of the purchasing power parity theory and making use again of our notation, we should conclude that (Equation 3):

$$P_1 = P_2^* \tag{3}$$

so that (Equation 4):

$$\frac{P_1}{\varepsilon P_2^*} = \frac{P_1}{P_2^*} \tag{4}$$

The ratios in Equation 4 represent two equivalent definitions of the purchasing power parity. In principle, we would like, in a perfect textbook situation, to work with both constant values and level variables. However, given the scope of our paper and considering our prices are time-indexed, what we tested for is the relative version of the purchasing power parity (Equation 5):

$$\frac{\Delta p_{1,t}}{\Delta p_{2,t}^*} = z_t \xrightarrow{yields} k \tag{5}$$

when $t \to \infty$. As the definition of relative PPP states, changes in national price levels are always equivalent to a constant value, or at least tend to such value (which does not always equal the nominal exchange rate, as we saw in the absolute version) in the medium or long run. Expressing the ratio in log-deviation to consider possible non linearities in the price index and assuming a constant rate of growth of price levels, we would get $\Delta(p) = \ln(p)$, which implies (Equation 6):

$$\ln\left(\frac{p_{1,T}}{p^*_{2,T}}\right) = \ln(p_{1,t}) - \ln(p^*_{2,t}) = z_t \tag{6}$$

Purchasing power parity can be modified introducing nonlinearities in its convergence rate. The commodity point theory by Heckscher states that a region might exist, delimited by a lower and an upper bound on z_t, where convergence is not existent, and the price differential does not show any central tendency which would make it go back to its attractor value x_t. Such area would be due to non-perfect arbitrage conditions, which Heckscher justified with the existence of either transportation costs or uncertainty. We would as such have (Equation 7):

$$z = \begin{cases} \dfrac{\Delta p^L_{1,t}}{\Delta p^{L*}_{2,t}} \to z_1 \; for \; t \to \infty \\ \dfrac{\Delta p^M_{1,t}}{\Delta p^{M*}_{2,t}} = k \; with \; k \in (z_1; z_2) \\ \dfrac{\Delta p^H_{1,t}}{\Delta p^{H*}_{2,t}} \to z_2 \; for \; t \to \infty \end{cases} \tag{7}$$

with $z_1 \neq z_2$, and where L, M and H identify the two possible states of convergence L and H, and one state of perhaps slower or nonexistent convergence, M. When the middle state is not present and the threshold is reduced to $z_1 = z_2 = z$, the commodity point theory simply collapses to a nonlinear, two regimes purchasing power parity model where price convergence might differ whether we are above or below the steady state value of x. Furthermore, to account for uncertainty and a non-perfect information, the model shown in Equation 7 can be empirically adjusted with a function that smooths transition between states. Note that in this theoretical section we have not mentioned that some long-run trends might be present in the price differentials. We will take such eventuality into account by detrending and demeaning z accordingly in the following sections.

3.1. Data

We considered an initial group of twenty-five commodity exporters. In search of a sufficiently long sample period in monthly frequencies,[7] our price differentials series were sourced from the International Monetary Fund (IMF) International Financial Statistics (IFS) database. The nominal CPI we employ were sourced following Bodart *et al.* (2015) country selection,[8] deflated by the US CPI, and finally demeaned and detrended according to the data generating process that was assumed based on the result of a battery of unit root tests. All the variables are expressed in logarithms, and as such adjusted for eventual non-stationarity in variance, interpretable in terms of elasticities for small deviations and suitable for the simple forecasting exercise we will carry on in the article. Thus, as we work in log-differences, the price differentials we employ are based on the following measure (Equation 8):

[7] Quarterly series were available. However, as past literature has highlighted, changes in economic decisions might happen at intervals much lower than months. Perhaps, the most frequently used frequency in past works related to the topic of exchange rate behavior has been the monthly one.
[8] Our selection basically entailed checking for all the available CPI series from the IMF IFS, which were employed by Bodart *et al.* (2015), conditional both on the availability of data, and on a suitable non-dollarized national currency (Ecuador was excluded from the analysis for such reason, while a CPI series for Dominica was not available at the IMF IFS at the time of this article).

$$z_t = \ln(CPI_i) - \ln(CPI_{US}) \tag{8}$$

where; the index "i" represents each of the sixteen countries that were available for the analysis. The countries we consider were selected conditional on the availability of data from the source we mentioned above and a minimum export weight of their leading exported commodity close to twenty percent. A list of the selected countries is reported in Tables 1 and 2.

Table 1. Country and leading commodity couples

Commodity	w_t	Weight
Cotton	**Benin**	61.00
	Mali	33.48
	Pakistan	20.52
Tobacco	**Malawi**	60.50
	Zimbabwe	19.53
Copper	**Zambia**	59.99
	Chile	30.79
Gold	**Mali**	54.05
	Burundi	35.45
	Ghana	28.56
Coffee	**Burundi**	50.98
	Ethiopia	46.43
	Uganda	36.87

Notes: Country and commodity couples according to Bodart *et al.* (2015). Highlighted in bold, the final sixteen countries group that made it to the final analysis.

Table 2. Country and leading commodity couples

Commodity	w_t	Weight
Uranium	**Niger**	41.73
	Benin	29.90
Cocoa	**Ivory Coast**	34.10
	Ghana	33.16
Aluminum	**Mozambique**	33.44
Soya	**Paraguay**	32.72
Fish	**Mauritania**	30.96
	Mozambique	19.87
Bananas	Dominica	29.20
	Ecuador	17.83
Tea	**Kenya**	21.20
Crustaceans	Mozambique	18.96

Notes: Country and commodity couples according to Bodart *et al.* (2015). Highlighted in bold, the final sixteen countries group that made it to the final analysis.

As we focus on the short-run adjustment properties of national price differentials, being the long run trend analysis not of our main concern given the scope of this paper, we will define a set of price differentials w_t as the detrended component of the price difference z_t (nominal domestic price level over US price level) given by (Equation 9):

$$z_t = \alpha + \beta t + w_t \tag{9}$$

The residuals from the above definition represent the error correction term we have used throughout the analysis. The formulation in Equation 9 has thus been used when no deterministic breaks could be found in the series, while the Zivot and Andrews's (1992) test equation has been employed when the former test gave us evidence of structural breaks. From the previous definition, the attractor/error correction term can be estimated as an ordinary least square (OLS) residual, while a constant had to be considered as data availability forced us to work with CPI, and not absolute prices. In this sense, and assuming a traditional auto-regressive process of order one [AR(1)], the standard work-horse formulation for convergence would be (Equation 10):

$$\Delta\widehat{w}_t - \lambda\widehat{w}_{t-1} + e_t \tag{10}$$

where; the error term is expected to be normally distributed with mean 0 and constant variance σ^2, while the parameter λ is expected to be bounded between 0 and -1 and represents the convergence speed. As demeaning and detrending were already considered in the first stage OLS, the model does not contain deterministic terms. Moreover, as the standard empirical formulation for the relative PPP does not contain any additional endogenous first differences, we decided to expand the models with additional lagged regressors based on a standard efficiency/parsimony trade-off analysis carried out by using maximum truncation lag rules together with classic information criteria (specifically, the Bayesian information criterion). The coefficient of the error correction term w_{t-1}[9] bears important information on the efficiency of arbitrage between markets, and we expect it to depend on the nature of the good(s) under consideration, the existence and type of transaction/transportation costs required to carry out the actual transaction, and other economic aspects of the activity such as the geographical distance between locations and the actual face value of the exchange rate.

Before getting to the estimates, we employ a series of tests to check for a unit root in a single piece-wise linear specification for every series. We check for three kinds of alternative hypothesis commonly accepted by literature: we employ the ADF (augmented Dickey Fuller test) to check for the benchmark linear stationary alternative, we employ the KPSS (Kwiatkowsky *et al.* (1992) stationarity test) to invert the hypotheses and consider the null of trend-stationarity, and finally we employ the ZA test (the Zivot and Andrews, (1992)) to evaluate the null of a linear unit root against the alternative (and rather realistic) hypothesis of a stationary alternative with one deterministic endogenous break, possibly in both the mean and the trend.[10] This latter test, in particular, presents the following test equations (Equations 11 and 12):

$$z_t = \mu + \Theta DU_t(T_b) + \beta t + \gamma\Theta DT_t(T_b) + \alpha z_{t-1} + \sum_{j=1}^{k} c_j \Delta z_{t-j} + w_t \tag{11}$$

$$\widehat{w}_t = z_t - \left[\hat{\mu} + \widehat{\Theta} DU_t(T_b) + \hat{\beta} t + \hat{\gamma}\widehat{\Theta} DT_t(T_b) + \hat{\alpha} z_{t-1} + \sum_{j=1}^{k} \hat{c}_j \Delta z_{t-j} + w_t\right] \tag{12}$$

where; DU_t is a dummy variable capturing a shift in the intercept at time T_b and DT_t is a trend break variable for a break occurring at time T_b. Furthermore, $DU_t = 1$ if $t > T_b$ and zero otherwise and $DT_t = t - T_b$ if $t > T_b$ and zero otherwise. The null hypothesis is rejected if the coefficient α is statistically significant. The test is a sequential procedure, with a ten percent trimming at both ends of the sample to ensure enough observations for each possible break model. Similar to a grid-search, it makes use of a different dummy for every conceivable break-date. The convenient aspect of the test is that the algorithmic procedure selects the value of the break where the calculated ADF statistic (the value of $\hat{\alpha}$) is at its minimum, thus representing the estimate that would be the most likely one to reject the unit root null.

---

[9] Or attractor, as it has been frequently called in univariate literature: see again Obstfeld and Taylor (1997).

[10] Notice that we might have also employed an additional test, the test by Lumsdaine and Papell (1997), which extends the test of Zivot and Andrews (1992) allowing for two breaks instead of one. However, to consider the issue of data mining and given the nature of the data, we have currently limited the analysis to just one break.

4. Econometric models

Let us start considering an implicit forward looking autoregressive specification of the kind (Equation 13):[11]

$$w_{t+s} = f\left(w_t, w_{t-d}, \dots, w_{t-(m-1)d}; \theta\right) + e_{t+s} \tag{13}$$

where; the error e_{t+s} in every given period $t, t+1, \dots, t+s$ is assumed to be white noise and uncorrelated with x_{t+s} and f represents a function which maps R^m to the real numbers' realm R, while θ represents the threshold value of the nonlinear specification. The general equation above represents variable w_t as a function of an embedded dimension (or model lags) m, several time delays d (delays in the transition between states), and the forecasting steps s. Considering an explicit linear representation (Equation 14) of the previous function:

$$w_{t+s} = \phi + \phi_0 w_t + \phi_1 w_{t-d}, \dots, \phi_m w_{t-(m-1)d} + e_{t+s} \tag{14}$$

and a convenient ADF reparameterization can be undertaken (Equation 15).

$$w_{t+s} = \phi + \lambda w_t + \zeta_1 \Delta w_{t-d}, \dots, \zeta_{m-1} \Delta w_{t-(m-2)d} + e_{t+s} \tag{15}$$

where; the new parameters would be: $\lambda = \phi_0 + \phi_1 + \cdots + \phi_m$ and $\zeta_i = -(\phi_{j+1} + \phi_{j+2} + \cdots + \phi_{m+j})$.

The theoretical advantage of such reparameterization is that the stability conditions for all the roots of the polynomial in the original equation. The fact that they should lie all inside the unit circle, can now be easily expressed just with: $-1 < \lambda < 0$. Furthermore, on the interpretation side, it allows us to consider a mean reverting process with a short run attractor w_t, whose coefficient will represent the speed of mean reversion of the price differentials (λ). Consider now an unrestricted, three regimes first order SETAR (Equation 16), which we express in a non-ADF form for simplicity):

$$w_t = \begin{cases} \rho_l w_{t-1} + u_t \; if \; w_{t-1} \leq \theta_l \\ \rho_m w_{t-1} + u_t \; if \; \theta_l \leq \; w_{t-1} \leq \theta_h \\ \rho_h w_{t-1} + u_t \; if \; w_{t-1} \geq \theta_l \end{cases} \tag{16}$$

The above model collapses $w_t = \rho w_{t-1} + u_t$ whenever $\theta_l = \theta_m = \theta_h = 0$, with $\rho_l = \rho_m = \rho_h$. In equivalent terms, the two regimes first order SETAR can be represented as (Equation 17):

$$w_t = \begin{cases} \rho_l w_{t-1} + u_t \; if \; w_{t-1} \leq \theta \\ \rho_h w_{t-1} + u_t \; if \; w_{t-1} \geq \theta \end{cases} \tag{17}$$

where; the number of relevant thresholds decreases to one.[12] All of the mentioned models can be easily extended by adding lags in each regimes (as well as intercepts and other deterministic components) and, most importantly, additional restrictions can be specified to allow for $\theta_l = \theta_h$ and $\rho_l = \rho_h$, imposing two external symmetric regimes.[13] A simplified, general TAR model with p lags, d delays and m regimes will thus take the form (Equation 18):

---

[11] For convenience we are dropping the hat notation from w.

[12] Notice that the models become an X-TAR (exogenous threshold TAR models) whenever 8 = mt—d. In a previous study from the author, the exogenous threshold is represented by the first order difference of the international price of petrol and is set to be a proxy for the volatility of the energy commodity market which would affect transportation costs in international trade flows and as such the propension to operate arbitrage whenever it is convenient. The transition variable will thus not per-se be the deviation from the equilibrium path of the price differentials but can be thought as one of its causes.

[13] Balke and Fomby (1997) called such model a two regimes "BAND TAR" model.

$$w_t = \begin{cases} \mu_1 + \rho_{1,1}w_{t-1} + \ \dots + \rho_{1,p1}w_{t-p1} + u_t \ if \ x_{t-d} \geq \theta_{m-1} \\ \mu_2 + \rho_{2,1}w_{t-1} + \ \dots + \rho_{2,p2}w_{t-p2} + u_t \ if \ \theta_{m-1} \geq \ x_{t-d} \geq \theta_{m-2} \\ \dots \\ \mu_m + \rho_{m,m2}wm + u_t \ if \ \theta_m \geq x_{t-d} \end{cases} \quad (18)$$

where; $x_{t-d} = w_{t-d}$ in the case of the self-exciting models and equivalent to any other selected exogenous source of variation in the case of models with an exogenous threshold.[14] In the following section, We will make use of the models discussed up until now together with a standard AR(1) model to check which one would fit better the data in an in-sample estimation exercise.

An additional nonlinear model, the LSTAR (a transition autoregressive model with a logistic function as a smoothing function), together with a simple AAR (additive autoregressive model), and a NNET (linear neural network model) have also been considered and used in this paper for an out-of-sample forecasting exercise on a forecast horizon of twenty periods, to check for consistency of the in-sample findings and evaluate the forecasting performance of the SETAR models comparing them to alternative nonlinear representations which furthermore do not support any specific economic theory. The LSTAR model can be specified as (Equation 19):

$$\begin{aligned} w_{t+s} = (\phi_1 + \phi_{10}w_t + \phi_{11}w_{t+d} + \phi_{1L} + w_{t-(L-1)d}(1 - G(\gamma, w_{t-d}) + \\ (\phi_2 + \phi_{20}w_t + \phi_{21}w_{t+d} + \phi_{2L} + w_{t-(H-1)d}(1 - G(\gamma, w_{t-d}) + e_{t+s} \end{aligned} \quad (19)$$

where; the two regimes result from a transition guided by function G, which would depend on the slope coefficient γ and the delayed threshold variable w_{t-d}. The reason behind the choice of a smoothing function is that, we remind, no transition between two states of the world can happen discretely with market segmentation and imperfect information: as such, a portion of those who could belong to either the lower or the higher regime would basically belong to none and be part of an “inaction band” where the series would most likely behave like a unit root. Symmetric regime applications for STAR functions have been widely used in past international economics and finance applications.[15] An additive (non-parametric) autoregressive model was instead specified as (Equation 20):

$$w_{t+s} = \mu + \sum_{i=1}^{m} s_i(w_{t-(i-1)d}) \quad (20)$$

where; d is once more the maximum delay embedded in the estimation and the s_i represent smooth functions in the form of cubic regression spines with a penalty. Finally, we employ a linear method to estimate a simple neural network model. Such model, with D hidden units, activation function g, embedded delay d, and maximum lag order of m can be represented by (Equation 21):

$$w_{t+s} = \beta_0 + \sum_{j=1}^{D} \beta_j g(\gamma_{oj} \sum_{i=1}^{m} \gamma_{ij} w_{t-(i-1)d}) \quad (21)$$

Equations 19, 20, and 21, as we already mentioned, have been employed only in Section 4.1 for out-of-sample forecasting performance comparisons.

4.1. Unit root analysis and demeaning

To avoid the issue of data mining and to accommodate a preliminary informal analysis, we decide to couple the standard ADF and KPSS tests with the ZA unit root test with one endogenous deterministic break. The advantage of this test over the seminal Perron (1989) test is represented

[14] Notice that when $x_{t-d} = \Delta w_{t-d}$, the SETAR model becomes an M-SETAR (momentum SETAR model).
[15] See for instance the seminal work of Taylor *et al.* (2001) in a univariate framework, but also the multivariate error correction estimates and subsequent Exponential STAR contribution by Baum *et al.* (2001).

by the fact that break dates are endogenously retrieved by a recursive conditional estimation, and the null hypothesis is built so that the final test statistic would be calculated where the model would be more likely to present a rejection of the null of unit root. The ZA test presents a trend-stationary alternative, specifically a single break in both the intercept and the trend in the alternative hypothesis. We use such test to sharpen the demeaning and detrending of z_t: while non-rejection of the null of unit root led us to simple demeaning and detrending, a positive result of the ZA pushed us to demean and detrend the differential so that the resulting w_t would be obtained from the residuals of the ZA test regression itself.

The unit root tests we performed are visible in Tables 3 and 4. The standard ADF test with truncation lag order set to Schwert maximum criterion failed to reject the unit root null in all cases but one, Mozambique, while consistently with that the KPSS shows a complete rejection coverage. To make sure the one break we choose through the endogenous selection would be at least statistically meaningful, we ran a series of BP (Bai and Perron (2003)) tests on a maximum of nine stochastic (slope) breaks. The BIC criteria for the alternative break specifications of the latter test would generally indicate a maximum of two possible breaks. In more than half of our sample of countries, the first or second BP chosen break dates were close to the ones indicated by the ZA test.

Table 3. Unit root tests

w_t	ADF	ZA	Break	ADF/ZA lags	KPSS	KPSS lags
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**
BEN	-2.5276	-12.4418***	25	6	4.2183***	3
MAW	-1.1491	-5.6229**	175	7	2.8411***	4
ZMB	-1.6447	-6.8332***	81	7	6.6401***	4
CHL	-1.8256	-4.6954	48	4	2.6045***	2
MAL	-2.2745	-7.0687***	80	7	1.9895***	4
BDI	-2.2745	-4.1121	191	7	8.8626***	4
GHA	-2.9801	-5.0313	182	7	8.6783***	4
ETH	-0.9927	-4.2515	244	7	7.0804***	4

Notes: Columns (II) and (III) indicates ADF and ZA t-statistics. Columns (V) and (VI) lag truncation orders are selected according to Schwert minimal criterion. Column (IV) is endogenously retrieved period of deterministic break. Column (VI) represents KPSS test statistic. Finally, ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

Table 4. Unit root tests

w_t	ADF	ZA	Break	ADF/ZA lags	KPSS	KPSS lags
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**
UGA	-1.515	-4.0858	105	6	6.8431***	3
NGR	-1.630	-7.3770***	168	7	1.5983***	4
CIV	-1.9145	-8.0522***	168	7	7.7910***	4
MOZ	-4.1421***	-6.3742***	18	6	6.3435***	3
PAR	0.33635	-5.5541**	113	7	8.2886***	4
MAU	-2.4677	-3.8544	228	7	7.5749***	4
DOM	-2.6165	-3.7657	231	7	7.9842***	4
KEN	-1.2847	-4.9459*	146	7	8.7764**	4

Notes: Columns (II) and (III) indicates ADF and ZA t-statistics. Columns (V) and (VI) lag truncation orders are selected according to Schwert minimal criterion. Column (IV) is endogenously retrieved period of deterministic break. Column (VI) represents KPSS test statistic. Finally, ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

4.2. Functional form tests

As a first result over the in-sample goodness of fit of the nonlinear models we considered, we report the estimated values and bootstrapped p-values of the Hansen (1999) F-test based on the null hypothesis of nonlinearity against the alternative of TAR behavior and on the null of two regimes TAR behavior against the alternative hypothesis of three regimes. It is perhaps not surprising to see, given the extended literature related to nonlinearity in major currencies/advanced countries, that the null of linearity stands completely unrejected only in three out of the sixteen countries we analyzed, namely Chile, Paraguay, and Zambia, while at least some degree of nonlinearity in the SETAR sense is detected in the remaining countries. This is visible in Tables 5 and 6, where the bootstrapped p-values of the F-test report rejection at least at the ten percent significance value in most of the countries for either the F-test alternative of a three regimes SETAR model or a two regimes SETAR model.

Although the F-tests results visible in Columns (II) to (V) of Tables 5 and 6 represent a valid way to infer on the functional form of the error correction mechanism, this ex-ante value is not indicative of whether we should choose a specific functional form over another. As a matter of fact, Tables 5 and 6 come with two additional columns, VI and VII, where the F-tests for the null of two regimes against an alternative specification of a BAND-TAR with two non-symmetric thresholds is evaluated for each country. The columns present mixed results: none of the two models clearly prevails over the other. In order to give better inference and be more conclusive about the in-sample goodness of fit of the models, we present in Section 4, the estimates, the in-sample mean absolute percentage error and the Bayesian information criterion for all the computed SETAR models plus the AR(1) benchmark.

Table 5. Hansen linearity test

w_t	F-test 1vs2	p-val.	F-test 1vs3	p-val.	F-test 2vs3	p-val.
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**
BEN	35.32827**	0.045	95.66599*	0.100	53.84535**	0.03
MAW	35.77584*	0.060	44.39750	0.254	7.966472	0.69
ZMB	31.11597	0.111	62.24627	0.170	28.7415	0.45
CHL	12.72579	0.365	19.02285	0.788	5.516975	0.94
MAL	49.60375**	0.019	91.43738**	0.023	36.57492**	0.12
BDI	21.06846	0.151	48.377380**	0.030	26.42547**	0.02
GHA	62.87736***	0.005	108.67622***	0.007	40.01488**	0.05
ETH	34.43561***	0.001	55.71441***	0.002	19.71789	0.205

Notes: Columns (II), (IV), and (VI) represent Hansen F-test values. Columns (III), (V), and (VII) illustrate bootstrapped p-value (currently 1000 iterations). Finally, ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

Table 6. Hansen linearity test II

w_t	F-test 1vs2	p-val.	F-test 1vs3	p-val.	F-test 2vs3	p-val.
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**
UGA	23.35154**	0.035	46.72797**	0.017	21.63437*	0.079
NGR	39.81548**	0.014	72.29901**	0.018	29.75963*	0.075
CIV	23.15175	0.142	59.48496*	0.083	34.49718*	0.090
MOZ	85.06133***	0.010	273.35555***	0.00	143.6875***	0.010
PAR	20.59989	0.213	40.30980	0.348	18.81873	0.370
MAU	48.56994***	0.000	84.19487***	0.00	31.48119***	0.000
DOM	26.41433**	0.022	48.20392**	0.025	20.54221	0.170
KEN	35.38025*	0.054	42.20667	0.288	6.312963	0.870

Notes: Columns (II), (IV), and (VI) represent Hansen F-test values. Columns (III), (V), and (VII) illustrate bootstrapped p-value (currently 1000 iterations). Finally, ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

5. Estimates and in-sample fitness

Estimates of the AR(1), two regimes SETAR, and unrestricted three regimes SETAR are reported in Tables 7 and 8. The AR(1), two regime and three regime SETAR would refer respectively to the classic PPP theory, an asymmetric variation of it[16], and the revised Heckscher commodity points theory with an inaction band. As we would expect, the magnitude of the various λ coefficients greatly vary not just across countries but also across price regimes.

[16] Depending on which between the home and the away (US) currency is currently devaluated with respect to the other. In a sense, it is just a less empirically valid rendition of the Heckscher commodity points theory.

We would generally find no absolute confirmation of the PPP nor of the Heckscher commodity points theory. In general, all countries would seem to find the attractor for the AR(1) process significant, exception done for Uganda, Chile and Ethiopia. In most cases, the SETAR three regimes model, which we considered as the ideal setup for representing the inaction band typically found in the advanced countries literature, showed significant unrestricted coefficients in the outer regimes and an inner unit root only in Ghana, Niger, Ivory Coast and Kenya. All in all, no specification appears to be significantly more suited than another, at least in terms of suggesting a prevailing economic theory.

Table 7. Nonlinear estimates I

	AR(1)	**SETAR 2 regimes**		**SETAR 3 regimes**			**X-MTAR OIL**	
w_t	**Lambda**	**LambdaL**	**LambdaH**	**LambdaL**	**LambdaM**	**LambdaH**	**LambdaL**	**LambdaH**
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**	**(VIII)**	**(IX)**
BEN	-0.222***	0.038	-0.161***	0.038	-0.191*	-0.093	-0.206	-0.205***
	('0.037)	(0.067)	(0.059)	(0.067)	(0.097)	(0.079)	(0.161)	(0.043)
MAW	-0.050***	-0.062***	-0.023	-0.046**	-0.081*	-0.037	-0.064	-0.032*
	(0.015)	(0.019)	(0.023)	(0.022)	(0.048)	(0.023)	(0.040)	(0.057)
ZMB	-0.037***	-0.022	-0.028	-0.017	-0.361***	-0.003	-0.055	-0.037
	(0.014)	(0.020)	(0.027)	(0.022)	(0.108)	(0.027)	(0.043)	(0.015)
CHL	-0.071	-0.040	-0.026	-0.040	0.001	-0.061	-0.065	-0.053**
	(0.022)	(0.034)	(0.029)	(0.034)	(0.051)	(0.038)	(0.049)	(0.022)
MAL	0.131***	-0.087**	-0.059	-0.078*	-0.208***	-0.033	-0.049	-0.151***
	(0.026)	(0.036)	(0.047)	(0.043)	(0.061)	(0.048)	(0.090)	(0.029)
BDI	-0.017**	-0.003	-0.026**	-0.006	0.007	-0.041**	-0.050	-0.017**
	(0.007)	(0.09)	(0.010)	(0.008)	(0.024)	(0.013)	(0.017)	(0.008)
GHA	-0.026***	-0.039***	-0.014	-0.039***	0.007	-0.018*	-0.015	-0.027***
	(0.007)	(0.09)	(0.009)	(0.009)	(0.031)	(0.009)	(0.025)	(0.007)
ETH	-0.002	-0.004	-0.023	-0.04	-0.008	0.001	-0.002	-0.012
	(0.003)	(0.005)	(0.005)	(0.005)	(0.012)	(0.006)	(0.004)	(0.008)

Notes: Column (II) represent AR(1) error correction term estimates, standard errors are shown in parentheses. Columns (III) and (IV) are error correction coefficient values for the self-exciting TAR with two regimes and one threshold specification. Columns (V), (VI) and (VII) are error correction coefficient values for the self-exciting TAR with 3 regimes and two thresholds specification. Columns (VIII) and (IX) represent the convergence values for a two regimes momentum threshold model with oil price variation as an exogenous threshold. ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

Table 8. Nonlinear estimates II

	AR(1)	**SETAR 2 regimes**		**SETAR 3 regimes**			**X-MTAR OIL**	
w_t	**Lambda**	**LambdaL**	**LambdaH**	**LambdaL**	**LambdaM**	**LambdaH**	**LambdaL**	**LambdaH**
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**	**(VIII)**	**(IX)**
UGA	-0.008	-0.000	-0.036**	-0.015	(0.016)	-0.036**	-0.013*	0.007
	(0.008)	(0.008)	(0.014)	(0.011)	(0.012)	(0.014)	(0.008)	(0.014)
NGR	-0.087***	-0.100***	-0.024	-0.105***	-0.046	-0.071**	-0.066***	-0.119***
	(0.017)	(0.024)	(0.036)	(0.024)	(0.066)	(0.034)	(0.023)	(0.041)
CIV	-0.099***	-0.133***	-0.045	-0.129***	0.048	-0.050*	-0.006	-0.069***
	(0.021)	(0.028)	(0.031)	(0.028)	(0.285)	(0.031)	(0.057)	(0.024)
MOZ	-0.071**	-0.119***	0.033	-0.118***	-0.172	-0.040	-0.100***	0.008
	(0.022)	(0.025)	(0.039)	(0.021)	(0.085)	(0.037)	(0.027)	(0.411)
PAR	-0.031***	-0.022	-0.037*	-0.014	-0.134**	-0.033	-0.024*	-0.032
	(0.012)	(0.524)	(0.021)	(0.015)	(0.055)	(0.022)	'(0.014)	(0.024)
MAU	-0.074***	-0.023	-0.155***	-0.018	0.006	-0.231***	-0.078***	0.027
	(0.021)	(0.027)	(0.034)	(0.031)	(0.038)	(0.049)	(0.023)	(0.050)
DOM	-0.078***	-0.050**	-0.054**	-0.036	-0.035	-0.044	0.060	-0.059**
	(0.017)	(0.024)	(0.025)	(0.024)	(0.075)	(0.026)	(0.047)	(0.018)
KEN	-0.064***	-0.034	-0.127***	-0.060**	0.023	-0.100***	-0.039	-0.063***
	(0.017)	(0.022)	(0.032)	(0.030)	(0.036)	(0.031)	(0.032)	(0.022)

Notes: Column (II) represent AR(1) error correction term estimates, standard errors are shown in parentheses. Columns (III) and (IV) are error correction coefficient values for the self-exciting TAR with two regimes and one threshold specification. Columns (V), (VI) and (VII) are error correction coefficient values for the self-exciting TAR with 3 regimes and two thresholds specification. Columns (VIII) and (IX) represent the convergence values for a two regimes momentum threshold model with oil price variation as an exogenous threshold. ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

In Table 9, we compare the implied half-lives of the deviations of the price differentials across regimes and models. As we would expect, the autoregressive estimates would indicate some of the half-lives ranging from three to more than six years. Strikingly, in nine of the countries we considered, deviations in the AR(1) specifications do not survive the one year. This is still high considering the volatility of the currency market, but perhaps lower than what implied by Rogoff (1996). The two-regime estimates do not generally confirm faster speed of mean reversion, and in some instances, quite puzzlingly, they show slower adjustment compared to the linear alternative, in both the low and the high regime.

Table 9. Half-lives estimates, AR(1) and BEST SETAR

	AR(1)		SETAR (BEST BIC)		
w_t	lambda	months	lambdaL	lambdaM	lambdaH
(I)	(II)	(III)	(IV)	(V)	(VI)
BEN	-0.222***	2.757	0	-	3.949***
MAW	-0.050***	13.555	10.830***	-	29.790
ZMB	-0.037***	18.284	31.159	-	24.407
CHL	-0.071	9.379	16.980	-	26.311
MAL	0.131***	4.948	7.615**	-	11.398
BDI	-0.017**	40.997	230.702	-	26.311**
GHA	-0.026***	26.595	17.424***	98.674	38.161*
ETH	-0.002	279.826	172.940	-	29.789
UGA	-0.008	88.381	Inf	-	18.905**
NGR	-0.087***	7.588	6.579***	-	28.533
CIV	-0.099***	6.683	4.857***	-	15.054
MOZ	-0.071**	9.363	5.520***	3.672	16.980
PAR	-0.031***	21.832	31.159	-	18.385*
MAU	-0.0735***	9.085	29.789	-	4.116***
DOM	-0.078***	8.572	13.513***	-	12.486***
KEN	-0.064***	10.515	20.038	-	5.103***

Notes: Half-life estimates of the attractors/error correction parameters, calculated according to $\lambda^T = (1-x)/(1-\lambda)$. Columns (IV), (V) and (VI) are estimations for the best BIC performing SETAR model. Finally, ***, **, and * represent significance levels at 1%, 5%, and 10%, respectively.

As we finally turn to the in-sample fitness of the models, we see that nonlinear modeling appears to outperform the benchmark linear specification in five out of the sixteen analyzed countries. We choose to evaluate the in-sample fitness using the Bayesian information criteria, whose results are visible in Table 10. Results did not differ that much when we used the Aikake information criterion and the mean absolute percentage errors, which are visible in Table 11.

Table 10. BIC results on model fitness

w_t	AR1	SETAR 2 regimes	SETAR 3 regimes	X-MTAR
(I)	(II)	(III)	(IV)	(V)
BEN	2175.675	-2173.325	-2144.747	-2180.051
MAW	-2598.134	-2577.289	-2565.858	-2579.004
ZMB	-1653.474	-1617.524	-1598.541	-1619.483
CHL	-1055.621	-1039.028	-1028.198	-1049.369
MAL	-2735.672	-2735.852	-2712.456	-2718.159
BDI	-3426.280	-3403.385	-3373.850	-3380.089
GHA	-3215.846	-3228.632	-3239.407	-3195.130
ETH	-3383.781	-3393.565	-3361.018	-3372.618
UGA	-2633.685	-2615.080	-2604.004	-2610.558
NGR	-3217.701	-3214.032	-3189.330	-3195.103
CIV	-3614.405	-3607.533	-3563.952	-3584.650
MOZ	-1902.421	-1923.465	-2005.764	-1933.375
PAR	-3036.307	-3027.375	-2992.899	-3000.880
MAU	-3236.322	-3234.195	-3224.534	-3222.511
DOM	-4169.419	-4148.434	-4127.868	-4156.693
KEN	-3173.778	-3139.624	-3126.026	-3147.530

Note: Columns (II) to (IV) represent Bayesian information criterion results for the in-sample estimates across all models.

Table 11. MAPE results on model fitness

w_t (I)	AR1 (II)	SETAR 2 regimes (III)	SETAR 3 regimes (IV)	X-MTAR (V)
BEN	2.492	2.695	2.915	3.130
MAW	2.088	2.303	2.130	2.952
ZMB	6.283	11.621	4.626	13.679
CHL	1.411	1.492	1.539	1.719
MAL	1.726	2.286	2.253	2.089
BDI	2.099	2.498	1.935	1.836
GHA	2.650	2.960	3.047	2.693
ETH	2.320	2.279	4.059	4.116
UGA	1.969	1.983	2.473	2.152
NGR	2.233	2.695	2.532	3.112
CIV	1.683	1.378	1.551	1.651
MOZ	3.983	10.429	4.715	6.899
PAR	2.542	2.841	2.668	2.215
MAU	1.820	2.883	1.802	2.146
DOM	1.544	1.914	2.007	1.703
KEN	2.461	2.472	2.657	2.575

Note: Columns (II) to (IV) show the mean average percentage error results for the in-sample estimates across all models.

In the upcoming Section 5.1 we will present our forecasting exercise and employ the MAPE once more to check how the alternative model specifications would perform outside of the sample.

5.1. Out-of-sample results

In our forecasting exercise, we apply a rolling window procedure over the first two-hundred and fifty-three consecutive periods (generally up to January 2011) for a forecast horizon of twenty months. As we have stated in the introduction of this work, and considering the mixed results obtained in the in-sample performance by the nonlinear models adopted, we choose to extend the exercise adding a second subgroup of models representing additional potential forms of nonlinearity in our price differentials, composed of an LSTAR with two regimes and a threshold delay of two and an amount of lags depending on Schwert minimal length criteria, and an AAR and LNN models with similar lag orders. As we performed the exercise on a group of countries with a nonstandard series length, we generally had enough observations to let the forecasting window remain fixed in relationship to the time-series span but had to modify it in the case of Mozambique, whose time series presented less than one hundred observations. Error forecast measurements for the forecasting exercise are available in Table 12.

The value for the SETAR belongs to the model which showed the lowest information criteria in the in-sample exercise. We remind the reader that the measure we are using is defined as (Equation 22):

$$MAPE = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{w_t - \widehat{w}_t}{w_t}\right| \tag{22}$$

where; w_t would be the actual value, $\widehat{w}_t$ the forecast, and $t \in (1;T)$ the forecasting horizon. Results of this exercise are quite striking: lowest percentage errors appear to be evenly distributed across all but the self-exciting TAR models, which perform very badly across all countries. The results for the percentage error, exception done for Ethiopia, appear to be relatively high and cast a shadow over asymmetrical TAR models as a possible alternative to univariate piecewise linear models in price differential forecasting. Best relative performance can finally be attributed to the AR(1) model and the NNET model. To check these findings, and considering the drawbacks of the percentage deviation measure we employed,[17] we calculated an additional measure, namely

[17] Among all, possible 0 divisions when real values are close to such limit, which happens to be the case in some instances of our differential measure.

the root mean square error (RMSE), where the divergence from the actual value would be standardized by the number of forecasting periods (Equation 23):

$$RMSE = \sqrt[2]{\frac{\sum_{t=1}^{T}(\widehat{w}_t - w_t)^2}{T}} \tag{23}$$

Results for this additional measure are reported in Table 13. Although with slight differences due to the approximation of the measures, the lowest root error appears to be evenly distributed across the various models, with the AR(1), the NNET, and the AAR generally contending the crown to best performer. Once more, the LSTAR model appears to contribute relatively less to the out-of-sample forecast performance, while the best BIC-selected SETAR model would never show a satisfying performance.

Table 12. Out-of-sample forecasting performance MAPE

w_t	AR1	SETAR	LSTAR	NNET	AAR
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**
BEN	141.332%	2189.900%	164.154%	151.376%	184.235%
MAW	64.34%	108.115%	46.716%	38.425%	40.706%
ZMB	25.959%	49.676%	31.861%	35.202%	25.345%
CHL	66.248%	530.981%	63.153%	61.412%	64.499%
MAL	103.397%	1470.881%	140.889%	109.904%	116.445%
BDI	309.489%	945.915%	294.758%	285.606%	290.494%
GHA	115.388%	128.156%	131.857%	134.527%	132.761%
ETH	7.027%	92.449%	7.308%	6.956%	7.524%
UGA	25.471%	109.531%	24.781%	23.473%	24.347%
NGR	225.867%	3591.119%	125.791%	150.204%	137.656%
CIV	92.460%	1077.125%	95.737%	93.782%	98.953%
MOZ	29.907%	361.428%	34.514%	19.998%	31.154%
PAR	41.697%	59.775%	62.103%	62.556%	46.763%
MAU	38.535%	129.570%	45.906%	40.335%	41.548%
DOM	339.129%	4920.56%	464.246%	436.971%	384.256%
KEN	95.0128%	736.287%	87.591%	105.982%	111.342%

Notes: Columns (II) to (VI) show the mean average percentage error for all the models employed in the analysis.

Table 13. Out-of-sample forecasting performance RMSE, rolling

w_t	AR1	SETAR	LSTAR	NNET	AAR
(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**
BEN	0.017	0.146	0.023	0.017	0.021
MAW	0.040	0.087	0.036	0.036	0.036
ZMB	0.060	0.156	0.067	0.069	0.060
CHL	0.004	0.030	0.004	0.004	0.003
MAL	0.010	0.091	0.011	0.010	0.010
BDI	0.015	0.074	0.015	0.014	0.015
GHA	0.017	0.042	0.020	0.019	0.019
ETH	0.045	0.377	0.045	0.045	0.046
UGA	0.021	0.107	0.021	0.021	0.021
NGR	0.011	0.097	0.011	0.011	0.011
CIV	0.010	0.060	0.011	0.010	0.011
MOZ	0.016	0.103	0.019	0.015	0.017
PAR	0.031	0.054	0.033	0.034	0.031
MAU	0.018	0.052	0.019	0.020	0.019
DOM	0.006	0.071	0.007	0.007	0.007
KEN	0.010	0.065	0.010	0.010	0.010

Notes: Column (II) to (VI) represent root mean square error for all the models employed in the analysis.

To ultimately cope with the somewhat more obscure interpretation of the RMSE, we also calculated the MAE (mean absolute error), which computes the average absolute difference between $\widehat{w}_t$ and w_t and allows proportional contribution of each error to the absolute measure. The MAE was calculated as (Equation 24):

$$MAE = \sum_{t=1}^{T} \left| \frac{\widehat{w}_t - w_t}{T} \right| \tag{24}$$

results are available in Table 14, and the conclusions we could draw from these results did not differ from those previously seen with other forecast deviation measures.

Table 14. Out-of-sample forecasting performance MAE

w_t	AR1	SETAR	LSTAR	NNET	AAR
(I)	(II)	(III)	(IV)	(V)	(VI)
BEN	0.011	0.146	0.013	0.010	0.013
MAW	0.032	0.074	0.025	0.023	0.024
ZMB	0.025	0.141	0.022	0.022	0.021
CHL	0.003	0.029	0.003	0.003	0.003
MAL	0.008	0.089	0.009	0.008	0.008
BDI	0.010	0.062	0.011	0.011	0.011
GHA	0.013	0.036	0.012	0.012	0.012
ETH	0.014	0.370	0.015	0.014	0.015
UGA	0.009	0.101	0.010	0.009	0.009
NGR	0.009	0.100	0.008	0.008	0.008
CIV	0.007	0.057	0.007	0.007	0.007
MOZ	0.011	0.087	0.013	0.010	0.010
PAR	0.011	0.044	0.011	0.012	0.011
MAU	0.007	0.039	0.008	0.008	0.008
DOM	0.005	0.070	0.005	0.005	0.005
KEN	0.006	0.053	0.006	0.006	0.006

Note: Columns (II) to (VI) represent mean absolute error for all the models employed in the analysis.

Table 15. Out-of-sample forecasting performance RMSE, recursive

w_t	AR(1)	SETAR	LSTAR	NNET	AAR
(I)	(II)	(III)	(IV)	(V)	(VI)
BEN	0.0393	0.1290	0.1061	0.0652	0.0571
MAW	0.2989	0.1817	0.2730	0.2826	0.2711
ZMB	0.3455	0.1085	0.3511	0.3320	0.3629
CHL	0.0145	0.0612	0.0263	0.0160	0.0154
MAL	0.0134	0.1119	0.0148	0.0201	0.0167
BDI	0.0810	0.0643	0.0663	0.0746	0.0722
GHA	0.0532	0.0374	0.0509	0.0774	0.0628
ETH	0.2095	0.2770	0.2102	0.2295	0.2045
UGA	0.1377	0.0347	0.1304	0.1312	0.1316
NGR	0.0222	0.1226	0.0284	0.0322	0.0277
CIV	0.0362	0.1246	0.0479	0.0425	0.0534
MOZ	0.0722	0.1975	0.1287	0.0409	0.0391
PAR	0.1834	0.0588	0.1936	0.2084	0.1823
MAU	0.0926	0.0221	0.1118	0.5696	0.1227
DOM	0.0632	-	0.0330	0.0305	0.0134
KEN	0.0901	0.1141	0.0878	0.0916	0.1067

Note: Column (II) to (VI) represent the root mean square error for all the models employed in the analysis.

One additional insight from this more direct measure of forecast error is that, perhaps surprisingly, the NNET and AAR models appear to statistically dominate the residual models even more, with eleven and twelve of the best relative results out of a total of sixteen countries, respectively.

To conclude our forecasting exercise, we present the recursive one-step-ahead forecast estimates of our models (Table 15) and the Diebold and Mariano tests (Table 16) evaluating any systematic difference between the linear autoregressive specification across the two different forecasting methods, aimed at helping us in the ex-post choice of the optimal forecast.

Table 16. Diebold Mariano tests

w_t	DM²	p-value	DM	p-value	AR(1) rolling	AR(1) recursive
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
BEN	18.509	0.0912	29.474	0.0133	0.0172	0.0393
MAW	81.317	0.0000	110.460	0.0000	0.0400	0.2989
ZMB	68.603	0.0000	175.900	0.0000	0.0604	0.3455
CHL	0.9647	0.3554	13.566	0.2021	0.0039	0.0145
MAL	-0.4503	0.6612	0.0152	0.9882	0.0096	0.0134
BDI	38.439	0.0027	58.825	0.0001	0.0148	0.0810
GHA	17.745	0.1036	21.085	0.0587	0.0171	0.0532
ETH	60.500	0.0001	43.984	0.0000	0.0445	0.2095
UGA	104.860	0.0000	114.13	0.0000	0.0208	0.1377
NGR	18.199	0.0961	18.154	0.0968	0.0112	0.0222
CIV	22.726	0.0441	31.910	0.0086	0.0101	0.0362
MOZ	68.556	0.0000	1.545.800	0.0000	0.0160	0.0722
PAR	83.821	0.0000	275.550	0.0000	0.0307	0.1834
MAU	66.221	0.0000	142.220	0.0000	0.0177	0.0926
DOM	182.240	0.0000	390.550	0.0000	0.0064	0.0632
KEN	58.573	0.0001	57.841	0.0001	0.0097	0.0901

Note: Column (II) and (IV) show test statistic based on squared and linear absolute difference loss function between the previous rolling and recursive raw forecasts.

Surprisingly, the rolling forecasts would suggest the SETAR models as an alternative as valid as (if not slightly better than) the linear specification for out-of-sample forecasting. This would represent weak evidence pushing our conclusion in favor of the Heckscher commodity points theory rather than the classic purchasing power parity theory.

However, we ultimately choose to specify a set of two Diebold-Mariano type tests to compare the autoregressive specification in both the rolling and the recursive estimates. In such instance, we have conclusive evidence over the better performance of the AR(1) in the rolling forecast exercise compared to the recursive estimates. It is safe to state that, when considering forecasting purposes, the simple and more parsimonious AR(1) model does not appear to perform any worse than any nonlinear model in the out-of-sample rolling window forecasting exercise. In theoretical terms, the AR(1) model still appears to outperform any possible theory-related nonlinear variant of the relative price adjustment in rolling window forecasting, but its interpretation as a long-run equilibrium definition is overshadowed by the results of the recursive estimates, which present a systematically lower RMSE and should theoretically guarantee relatively lower error forecasts given that the procedure fixes the beginning of the forecast window at the very beginning of the sample.

6. Conclusions

In this work, we presented an analysis of sixteen commodity countries ‘exchange rate movements in relationship to the US dollar. In-sample fitness of sixteen US CPI-relative national price differentials is evaluated and modelled using a set of threshold nonlinear models, while out-of-sample forecast accuracy is evaluated based on rolling and fixed window forecasts through calculation of mean absolute percentage errors and extended to three additional models, namely a logistic smooth transition regression, an additive nonlinear autoregressive model, and a simple Neural Network model.

Our preliminary results confirm presence of a form of TAR nonlinearity in most of the countries analyzed. SETAR models tend to have quite poor relative performance in the rolling window exercise, both when compared to alternative nonlinear specifications and to the benchmark linear model but perform at least as well as the linear specification in the recursive exercise, pushing our conclusions towards the acceptance of the commodity points theory. Our results confirm the presence, in the spirit of Obstfeld and Taylor (1997) review of the Heckscher commodity points theory, of band convergence with an inner unit root in seven of the sixteen countries considered once the final recursive forecasting estimates are accounted for.

As a final policy advice to commodity-exporting countries, we find no apparent reason to suggest commodity export price pegging as a generalized policy perhaps given the presence of strong nominal frictions dampening relative price adjustment. As we compare it to the results of the most recent works, such as Bodart *et al.* (2012) and Coudert *et al.* (2015), we can conclude that, although the speed of mean reversion might be indeed lower than what such researchers would suggest, it is indeed true that, within the limits of a national export structure focused on commodities, long-run convergence and possible forecastability through new nonlinear techniques might still be a feasible and interesting possibility.

References

Allen, D. E., McAleer, M., Peiris, S., and Singh, A. K., 2016. Nonlinear time series and neural-network models of exchange rates between the US Dollar and major currencies. *Risks*, 4(1), 7. https://doi.org/10.3390/risks4010007

Bai, J., and Perron, P., 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), pp. 1-22. https://doi.org/10.1002/jae.659

Balke, N., and Fomby. T., 1997. Threshold cointegration. *International Economics Review*, 38(3), pp. 627-645. https://doi.org/10.2307/2527284

Baum, C. F., Barkoulas, J. T., and Caglayan, M., 2001. Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. *Journal of International Money and Finance*, 20(3), pp. 379-399. https://doi.org/10.1016/S0261-5606(00)00043-7

Bodart, V., Candelon, B., and Carpantier, J. F., 2012. Real exchange rates in commodity producing countries: a reappraisal. *Journal of International Money and Finance*, 31(6), pp. 1482-1502. https://doi.org/10.1016/j.jimonfin.2012.02.012

Bodart, V., Candelon, B., and Carpantier, J. F., 2015. Real exchange rates, commodity prices and structural factors in developing countries. *Journal of International Money and Finance*, 51(C), pp. 264-284. https://doi.org/10.1016/j.jimonfin.2014.11.021

Cashin, P., Cespedes, L. F., and Sahay, R., 2004. Commodity currencies and the real exchange rate. *Journal of Development Economics*, 75(1), pp. 239-268. https://doi.org/10.1016/j.jdeveco.2003.08.005

Chen, Y., and Rogoff, K., 2003. Commodity currencies. *Journal of International economics,* 60(1), pp. 133-160. https://doi.org/10.1016/S0022-1996(02)00072-7

Chen, Y., Rogoff, K., and Rossi, B., 2010. Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics*, 125(3), pp. 1145-1194. https://doi.org/10.1162/qjec.2010.125.3.1145

Coudert, V., Couharde, C., and Mignon, V., 2015. On the impact of volatility of the real exchange rate/terms of trade nexus: revising commodity currencies. *Journal of International Money and Finance*, 58, pp. 110-127. https://doi.org/10.1016/j.jimonfin.2015.08.007

Ferraro, D., Rogoff, K., and Rossi, B., 2015. Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance*, 54, pp. 116-141. https://doi.org/10.1016/j.jimonfin.2015.03.001

Frankel, J. A., 2005. Peg the export price index: a proposed monetary regime for small countries. *Journal of Policy Modeling*, 27(4), pp. 495-508. https://doi.org/10.1016/j.jpolmod.2005.04.013

Frankel, J. and Saiki, A., 2002). A proposal to anchor monetary policy by the price of the export commodity. *Journal of Economic integration*, 17(3), pp. 417-448. https://doi.org/10.11130/jei.2002.17.3.417

Hansen, B., 1999. Testing for linearity. *Journal of Economic Surveys*, 13(5), pp. 551-576. https://doi.org/10.1111/1467-6419.00098

Harvey, D. I., Kellard, N. M., Madsen, J. B., and Wohar, M. E., 2010. The Prebisch-Singer hypothesis: four centuries of evidence. *The Review of Economics and Statistics*, 92(2), pp. 367-377. https://doi.org/10.1162/rest.2010.12184

Kwiatkowsky, D., Phillips, P. C. B., Schmidt, P., and Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), pp. 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lumsdaine, R. L. and Papell, D. H., 1997. Multiple trend breaks and the unit-root hypothesis. *The Review of Economics and Statistics*, 79(2), pp. 212-218. https://doi.org/10.1162/003465397556791

Meese, R., and Rogoff, K., 1983. Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics*, 14(1-2), pp. 3-24. https://doi.org/10.1016/0022-1996(83)90017-X

Michael, P. A., Noobay, A. R., and Peel, D. A., 1997. Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy*, 105(4), pp. 862-879. https://doi.org/10.1086/262096

Obstfeld, M. and Taylor A. M., 1997. Nonlinear aspects of goods market arbitrage adjustment: Heckscher's commodity points revisited. *Journal of the Japanese and International Economies*, 11(4), pp. 441-479. https://doi.org/10.1006/jjie.1997.0385

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), pp. 1361-1401. https://doi.org/10.2307/1913712

Ricci, L. A., Milesi-Ferretti, G. M., and Lee, J., 2013 Real exchange rates and fundamentals: a cross-country perspective. *Journal of Money, Credit and Banking,* 45(5), pp. 845-865. https://doi.org/10.1111/jmcb.12027

Rogoff, K., 1996. The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2), pp. 647-668.

Rossi, B., 2013. Exchange rate predictability. *Journal of Economic Literature*, 51(4), pp. 1063-1119. https://doi.org/10.1257/jel.51.4.1063

Taylor, M. P., Peel, D. A., and Sarno, L., 2001. Nonlinear mean reversion in real exchange rates: toward a solution to the purchasing power parity puzzles. *International Economic Review*, 42(4), pp. 1015-1042. https://doi.org/10.1111/1468-2354.00144

Zivot, E. and Andrews, D. W. K., 1992. Further evidence on the great crash, the oil-price shock and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 20(1), pp. 25-44. https://doi.org/10.1198/073500102753410372

Appendix

This appendix reports the descriptive statistics (Table 17), a graphical rendition of the z_t time series for Kenya, together with the detrended series w_t and the out-of-sample graphic output of the forecasts (Figure 1 to 3).

Table 17. Descriptive statistics and demeaned price differentials

w_t	Min.	1st Q.	Median	Mean	3rd Q.	Max.
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
BEN	-0.389	-0.013	0.005	0.000	0.024	0.078
MAW	-0.840	-0.095	-0.007	0.000	0.109	0.367
ZMB	-1.982	-0.194	0.101	0.000	0.281	0.469
CHL	-0.030	-0.016	0.003	0.000	0.012	0.058
MAL	-0.281	-0.023	0.005	0.000	0.025	0.102
BDI	-0.393	-0.078	0.026	0.000	0.095	0.374
GHA	-0.682	-0.110	0.019	0.000	0.130	0.492
ETH	-0.540	-0.205	0.033	0.000	0.193	0.488
UGA	-0.170	-0.110	0.020	0.000	0.094	0.139
NGR	-0.367	-0.025	0.007	0.000	0.035	0.194
CIV	-0.295	-0.017	0.001	0.000	0.017	0.100
MOZ	-0.526	-0.048	0.024	0.000	0.062	0.119
PAR	-0.528	'-0.107	0.050	0.000	0.100	0.238
MAU	-0.095	-0.024	0.010	0.000	0.025	0.111
DOM	-0.078	-0.016	0.002	0.000	0.017	0.068
KEN	-0.386	-0.042	0.011	0.000	0.048	0.231

Note: Descriptive statistics for every series. The variables have been demeaned (see Column (V)) and detrended.

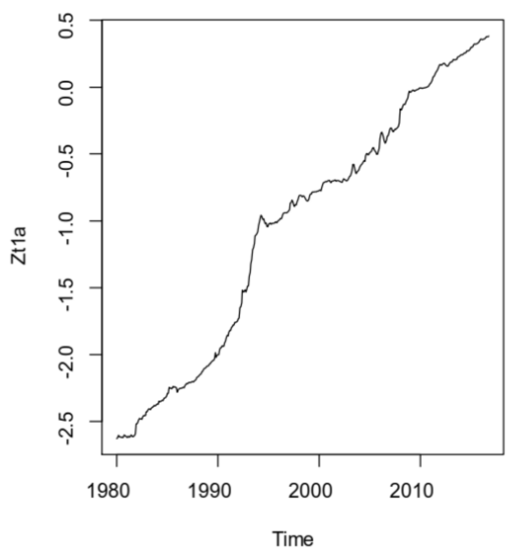

Figure 1. Kenya's CPI differential relative to the USA (z_t)

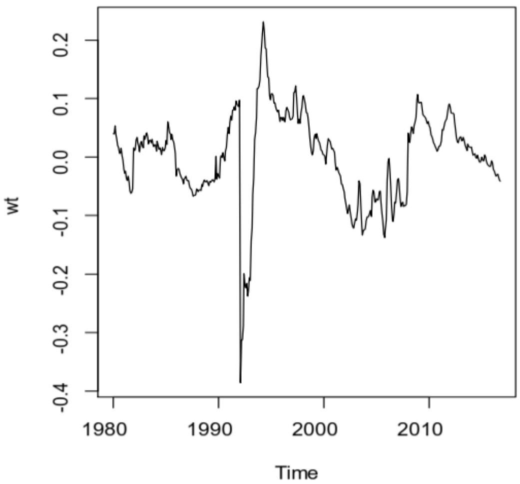

Figure 2. Kenya's demeaned and detrended z_t (in other words,(w_t))

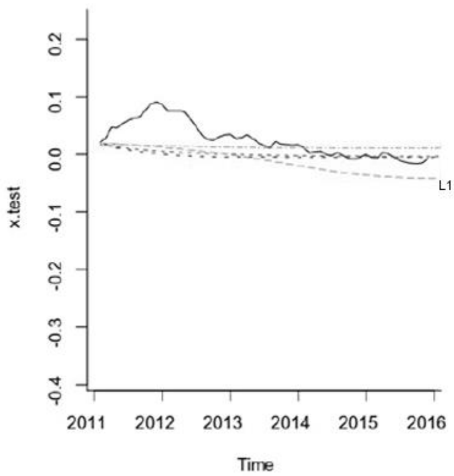

Figure 3. Kenya's rolling forecasts: SETAR based forecasts shown in L1