BANK OF FINLAND DISCUSSION PAPERS

13/2000

Biing-Shen Kuo – Anne Mikkola

Research Department
3.10.2000

Forecasting the Real US/DEM Exchange Rate: TAR vs. AR

SUOMEN PANKIN KESKUSTELUALOITTEITA • FINLANDS BANKS DISKUSSIONSUNDERLAG

Suomen Pankki
Bank of Finland
P.O.Box 160, FIN-00101 HELSINKI, Finland
☎ + 358 9 1831

Biing-Shen Kuo* – Anne Mikkola**

Research Department
3.10.2000

Forecasting the Real US/DEM Exchange Rate: TAR vs. AR

* National Chengchi University, Graduate Institute of International Trade, Taipei 116, Taiwan. E-mail: bsku@nccu.edu.tw.
** Bank of Finland, Research Department and University of Helsinki, Department of Economics. E-mail: anne.mikkola@helsinki.fi.

We would like to thank Juha Tarkka, Jouko Vilmunen and Matti Virén for helpful insights and discussions along the way.

ISBN 951-686-675-1
ISSN 0785-3572
(print)

ISBN 951-686-676-X
ISSN 1456-6184
(online)

Suomen Pankin monistuskeskus
Helsinki 2000

Forecasting the Real US/DEM Exchange Rate: TAR vs. AR

Bank of Finland Discussion Papers 13/2000

Biing-Shen Kuo – Anne Mikkola
Research Department

Abstract

The out-of-sample forecasting performances of two univariate time series presentations for the USD/DEM real exchange rate are compared using quarterly data for the period 1957Q1–1998Q4. The linear AR process is frequently fitted to real exchange rate series because it is sufficient for capturing the reported slow mean reversion in real exchange rates and it has some predictive ability for the long run. A simple nonlinear alternative, the threshold autoregressive (TAR) model, allows for the possibility that there is a band of slow or no convergence around the purchasing power parity level in the real exchange rate, due to transportation costs or other market frictions that create barriers to arbitrage. The TAR model is theoretically and empirically appealing, and it has been fitted to real exchange rates in many recent papers. However, the ultimate test of its usefulness is its out-of-sample forecasting accuracy. We compare the TAR model to its simple linear AR alternative in terms of out-of-sample forecast accuracy. Preliminary results using the RMSE criterion indicate that TAR forecasts are more sensitive to the estimation period and that they involve considerably more uncertainty at long horizons, as compared with the simple AR model.

Key words: real exchange rate, TAR model, forecast accuracy

TAR- ja AR-mallin tarkkuus Yhdysvaltain dollarin ja Saksan markan välisen reaalisen valuuttakurssin ennustamisessa

Suomen Pankin keskustelualoitteita 13/2000

Biing-Shen Kuo – Anne Mikkola
Tutkimusosasto

Tiivistelmä

Tässä tutkimuksessa verrataan kahden erilaisen reaalista valuuttakurssia kuvaavan aikasarjamallin ennustetarkkuutta otoksen ulkopuolisessa aineistossa. Tarkastelut perustuvat Yhdysvaltain dollarin ja Saksan markan väliseen reaaliseen valuuttakurssiin vuoden 1957 ensimmäisestä neljänneksestä vuoden 1998 viimeiseen neljännekseen. Usein reaalista valuuttakurssia mallinnetaan lineaarisella AR-mallilla, joka riittää kuvattaessa reaalisten valuuttakurssien taipumusta palata hitaasti kohti tasapainoarvojaan. Samalla sen avulla voidaan jossain määrin ennustaa reaalista kurssia pitkällä aikavälillä. Yksinkertainen AR-mallin epälineaarinen vaihtoehto on kynnysauto-regressiivinen (TAR) malli. Sen avulla voidaan mallintaa kuljetuskustannusten tai muiden markkinajäykkyyksien luomat vaihdannan esteet, joiden vuoksi reaalisen valuuttakurssin ostovoimapariteetin ympärillä on ”putki”, jonka sisällä kurssit palaavat hitaasti tai eivät lainkaan kohti tasapainoa. TAR-malli on sekä teoreettisesti että empiirisesti houkutteleva vaihtoehto, ja sitä on käytetty reaalisten kurssien mallintamiseen useissa viimeaikaisissa tutkimuksissa. Sen käyttökelpoisuutta voidaan testata parhaiten tarkastelemalla sen ennustetarkkuutta. Tässä tutkimuksessa TAR-mallin ja sen yksinkertaisen lineaarisen AR-vaihtoehdon ennustetarkkuutta verrataan otoksen ulkopuolisessa aineistossa. Alustavat, virheiden neliösummakriteeriin perustuvat tulokset osoittavat, että TAR-ennusteet ovat herkempiä valitun estimointiajanjakson suhteen ja että niihin pitkinä ennustejaksoina sisältyy huomattavasti enemmän epävarmuutta kuin yksinkertaiseen AR-malliin.

Asiasanat: reaalinen valuuttakurssi, kynnysmalli, ennustetarkkuus

Contents

1 Introduction

Our aim is to compare the forecasting ability of the standard autoregressive model and the so called threshold autoregressive (TAR) model in predicting real exchange rates. These are the most common univariate models used in describing the real exchange rate behaviour.

There is a large literature debating on whether real exchange rates are stationary or non-stationary. In finite samples, it is ultimately not possible to find the truth, since any process with a unit root can be approximated infinitely closely by a stationary process. Based largely on this literature, controversy remains on the usefulness of purchasing power parity in understanding the long run behavior of real exchange rates. Some authors view the time series evidence as casting doubt on the usefulness of PPP (e.g. O'Connell, 1998b; Engel 1996), while others do not see the evidence strong enough to discard the PPP even as an empirically relevant starting point (e.g. Lothian, 1997; Kuo-Mikkola, 1999a,b).

For long-run forecasting purposes the stationarity or non-stationarity of the specification is crucial. To the extent that real exchange rates are non-stationary, they cannot be predicted in the long run. In our exersice, we can expect to be able to forecast the real exchange rates only to the extent that the stationary element is present and captured by these common autoregressive specifications. More specifically, our interest is in judging the importance of TAR-type non-linearity by doing the out-of-sample forecast comparison relative to the simple linear specification.

The simple AR-model is found to fit quite well in the real exchange rates, and it has some long-run predictive power relative to a unit root processes (see Lothian-Taylor, 1996; Lothian, 1998; Kuo-Mikkola, 1999a). It is sufficient to capture the mean reverting nature of real exchange rates while at the same time allowing for fluctuations that persist over several years. Recently, TAR models have been fitted to real exchange rates or relative goods prices (e.g. Obstfeld-Taylor, 1997; Coakley-Fuertes, 1998; O'Connell-Wei, 1997). TAR model is an attractive alternative to AR-models, since it appears to reconcile the observation of long run stationarity with the periods where the RERs appears statistically non-stationary. The TAR model estimates a band around the long-run RER inside which the RER is nonstationary or follows a different stationary process from what it follows outside the band. Intuitively and theoreticallly this makes sense by allowing for a band of slow or no mean reversion around the purchasing power parity level, while the PPP deviations start to get eliminated more quickly as they become larger. The idea is that with small deviations the arbitrage is not active due to e.g. transportation or other transaction costs.

Ultimately economists want to develop models that would help us understand not only the past but also the future behaviour of real exchange rates. In order to compare the alternative models the ultimate test is to see how they perform in out-of-sample forecasting. Forecasting is of interest to many applied economists as such, and one would want to know how well the alternatives do in this respect. On some occations one may well be content to estimate a simpler AR model, even if the TAR model were to provide a nicer economic interpretation, as long as the prediction is good enough.

2 Data

The real exchange rate between the US and Germany is constructed from the consumer price index series and the exchange rate series over 1957:1-1998:4. The real exchange rate for the US at time t, y_t, is

$$y_t = e_t - p_{USt} + p_{GERt},$$

where p_{USt} and p_{GERt} are the CPIs for the US and Germany at time t. e_t is the price of German Mark in terms of the US dollar at time t. All variables are in logarithms.[1].

The persistence in the real exchange rates as measured by the sum of AR-coefficients is considerably less in annual than in monthly or quarterly data. Generally, the persistence is the lower the lower the frequency of the data. It is this lack of persistency or mean-reversion that we want to make use of in the out-of-sample forecasting exercise. Therefore, we want to use as low frequency data as possible and chose quarterly rather than monthly data in this exercise. In other words, we focus on forecastablity due to the long-run behavior where arbitrage leads to convergence in the purchasing powers of the two currencies.

3 The competing models

Simple AR model is often used to study the potential convergence of prices and exchange rates towards the purchasing power parity levels. The AR(p) specification for the real exchange rate, y_t, is

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_q y_{t-p} + e_t \tag{1}$$

Convergence speed is then interpreted to measure market integration or efficiency of arbitrage. However, if the implied market frictions are transportation costs, tariffs, non-tariff barriers, menu costs or even pricing to market behavior, then there can be deviations from the law of one price without arbitrage taking place. These frictions create a band of no convergence around the real exchange rate implied by PPP. Within the band, the nominal exchange rate can move around without the relative domestic prices responding. This was pointed out by Hechscher in 1916 and recently elaborated by Obstfeld and Taylor (1997).

A simple way to introduce the idea of the band of slow or no convergence is to modify the AR model into the TAR model: one linear autoregressive process is replaced by two linear autoregressions, thus leading to a non-linear presentation for the real exchange rates in time. In the TAR representation, the observations are split into two regimes, the inner regime where there is sluggish or no adjustment and the outer regime where large deviations from PPP create profitable arbitrage and mean reversion towards equilibrium takes place.

We fit the so called Equilibrium Tar (EQ-TAR) model to the data. The mean real exchange rate in the sample period, $\overline{y}$, is taken as the rough approximation

[1] Data is extracted from the IMF publication, *International Financial Statistics*. The consumer price index series (IFS line 64) is used as the measure of prices, and the price of U.S. dollars in DEM (IFS line zf) as the raw exchange rate series. This series is an average over the period number.

of the equilibrium real exchange rate. The real exchange rate's behavior today is determined by one of the two linear AR-processes. y_t is hypothesized to follow a more sluggish inner regime AR process if the the previous period's real rate is no further than c from the equilibrium and the faster converging outer regime process otherwise. The threshold value, c, is thus used to split the sample into two regimes. The EQ-TAR process can be written as follows.

$$\Delta y_t = \begin{cases} \lambda_1(y_{t-1} - \overline{y}) + \cdots + \lambda_p(y_{t-p} - \overline{y}) + e_t & q_{t-d} > c \\ \beta_1(y_{t-1} - \overline{y}) + ... + \beta_q(y_{t-q} - \overline{y}) + e_t, & -c \leq q_{t-d} \leq c \\ \lambda_1(y_{t-1} - \overline{y}) + \cdots + \lambda_p(y_{t-p} - \overline{y}) + e_t & q_{t-d} < -c \end{cases} \tag{2}$$

where q_{t-d} is the threshold variable, typically the lagged real exchange rate. We set $d = 1$ and $q_{t-d} = y_{t-1} - \overline{y}$. Thus, the position of the previous period's real exchange rate relative to the sample mean determines the regime in which the real exchange rate is today. Notice that in the EQ-TAR specification the threshold, c, is only used to split the sample but it does not itself appear in the two time series processes. Consequently, the process converges to the overall mean in both regimes. The process is stationary overall if the outer band dynamics are stationary: the process always reverts to the inner band in this case.

This model will be estimated as follows:

$$\Delta y_t = \begin{cases} \lambda_0 + \lambda_1 y_{t-1} + \cdots + \lambda_p y_{t-p} + e_t, & q_{t-1} > c \text{ or } < -c\,, \\ \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_q y_{t-q} + e_t, & -c \leq q_{t-1} \leq c \end{cases} \tag{3}$$

An alternative frequently used TAR specification is the so called BAND-TAR model.

$$\Delta y_t = \begin{cases} \lambda_1(y_{t-1} - \overline{y} - c) + ... + \lambda_p(y_{t-p} - \overline{y} - c) + e_t, & q_{t-d} > c \\ \beta_0 + \beta_1(y_{t-1} - \overline{y}) + ... + \beta_q(y_{t-q} - \overline{y}) + e_t, & -c \leq q_{t-d} \leq c \\ \lambda_1(y_{t-1} - \overline{y} + c) + ... + \lambda_p(y_{t-p} - \overline{y} + c) + e_t, & q_{t-d} < -c \end{cases} \tag{4}$$

As opposed to our EQ-TAR specification, the convergence from the outer regimes is to the bands rather than the mean. O'Connell (1998a) points out that the EQ-TAR specification is likely to provide a reasonable approximation to other candidate TAR processes, such as the BAND-TAR. Examples of papers fitting a BAND-TAR process to real exchange rates are Coakley and Fuertes (1998) and Obstfeld and Taylor (1997), while O'Connell (1998a) fits an EQ-TAR model.

For out-of-sample forecasting purposes it may be beneficial to have a specification like EQ-TAR where both regimes converge towards the "PPP" value. This is more so because both the thresholds and the equilibrium value are estimated and involve themselves a lot of uncertainty. The EQ-TAR model is estimated in the form of (3), which allows for free estimation of the constant and thus the point of convergence for the two regimes separately. Restrictions that could be imposed would be to set $\beta_0 = \beta_1 = \cdots = \beta_q = 0$ or to estimate the EQ-TAR without a constant as in (2) with the convergence strictly forced to the sample mean rather than allowing for its estimation. Given that the equilibrium value is an imprecise approximation and our focus is in out-of-sample forecasting, we want to impose as few restrictions as possible ex ante.

Why is the TAR-model considered as a plausible specification for real exchange rates?

Firstly, the TAR specification is a simple way to modify the standard AR-model to account for the effect of trading frictions. Secondly, there is some empirical evidence that the convergence is indeed faster the farther away the real exchange rate is from the PPP value. Michael, Nobay, Peel (1997) find support for this in annual data extending over two centuries and Obstfeld and Taylor (1997) in monthly data over 1980-95 as do Coakley and Fuertes (1998) on monthly post-Bretton Woods data. O'Connell's (1998a) results are more split. TAR specifications tend to indicate overall faster mean reversion than the AR models, which would help to explain the puzzling slow convergence of the PPP studies. Thirdly, the existence of time periods of slow or no convergence alongside with periods of faster convergence would explain some of the controversy concerning the empirical evidence on the stationarity of the real exchange rates. Alternating periods of convergence and non-convergence would surely make the statistical testing more difficult. Indeed, O'Connell (1998a) shows that a band of no arbitrage created by transport costs of only 6-10% can explain the failure of DF tests of rejecting the null of non-stationarity in the real exchange rates.

4 Estimation

4.1 The AR model

The baseline AR(p) model in (1) is estimated with two lags. Both Akaike and Schwarz criterion choose $p = 2$ when the lag order was varied from one to four. Out-of-sample forecasting accuracy measured by the RMSEs is also not sensitive to the order of the AR process. $AR(2)$ - appears sufficient and is doing slightly better than the higher order processes. Therefore, we fix the number of lags to two – also for the TAR-processes in (3), i.e. p and q are set equal to two in the following estimations. Estimation results are reported in table 1. The sum of the AR coefficients is negative, -0.03, implying slow mean reversion as expected. This is the benchmark model to which the out-of-sample forecast performance of the TAR model will be compared.

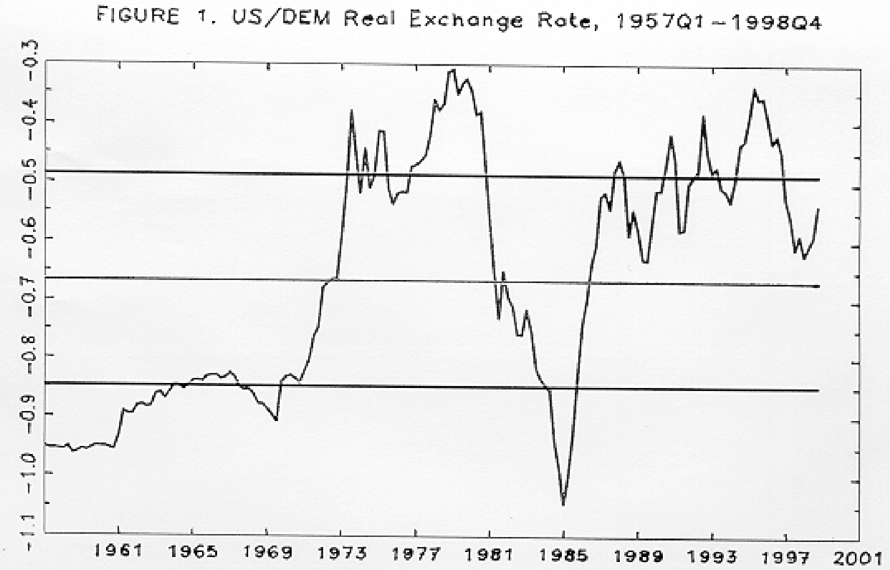
4.2 The EQ-TAR model

To estimate the threshold regression model, we utilize Bruce Hansen's (2000, hereafter BH) code. BH's model has one threshold which is used to split the sample into two groups - or regimes. Instead of one threshold we have two (c and $-c$), but essentially the model is the same, since only one c and two separate regressions for the two regimes are estimated. Consequently, the programs require only minor modifications.

The parameters of (3), $(\lambda_0, ..., \lambda_p, \beta_0, ..., \beta_q, c)$, are estimated by least squares as follows. To estimate c, a grid search is done over possible values of the threshold variable. Only 70% of the observations are used to choose c from to ensure that the model is well identified for all threshold candidates (see Hansen, 1996b). Thus, we trim the bottom and top 7.5% of the q_ts along with 15% around the mean. For each c, the observations are sorted into the outer and inner regimes, and both are estimated by least squares. The sum of squared errors, $SSE(\lambda_0, ..., \lambda_p, \beta_0, ..., \beta_q, c)$ is calculated and the c minimizing the SSE is selected. The null of no heteroscedasticity cannot be rejected for the estimated TAR model[2].

The full sample estimation results for the TAR model are presented in table 1. Figure 1 depicts visually the position of the estimated thresholds ($c = 0.18$ and $-c = -0.18$) as deviations from the full sample mean of the time series (the middle line at -0.667). The observations falling within the two outer lines are the ones that are estimated to follow the inner regime process, while the remaining observations follow the outer regime process. Slightly more than half of the observations fall into the outer regime. As can be read from table 1, the outer regime indeed is mean reverting with the sum of AR coefficients being negative, -0.05. The implied outer regime convergence is somewhat faster than implied by the linear AR process. The inner regime process does not show any convergence. Indeed, we might want to restrict the inner regime to be nonstationary. The full sample results are thus not contradicting our initial hypotheses of faster outer regime convergence. In the forecasting exercise, the TAR and AR processes are estimated for shorter training periods since some of the observations are needed to do the out-of-sample comparison. Consequently, the results in table 1 are used only to give an idea of the general fit of the two models.

FIGURE 1. US/DEM Real Exchange Rate, 1957Q1–1998Q4



[2] LM test is used to test for heteroscedasticity. It is performed by regressing the squared residuals from (3) against all the independent variables as follows. $e_t^2 = \gamma_0 + \gamma_1 \, y_{t-1} + \gamma_2 \, y_{t-2} + \nu_t$. Under the null of no heteroscedasticity, $\gamma_1 = \gamma_2 = 0$, and the LM test statistic $(T(1 - \sum v_t^2 / \sum e_t^2)$, converges to a $\chi^2(2)$.

TABLE 1
Full sample estimation results for the EQ-TAR and AR(2) models

	EQ-TAR			AR
	Outer regime	**Inner regime**	**Full model**	
const	-0.04	0.04		-0.02
	(0.01)	(0.02)		(0.01)
coef1	0.24	0.37		0.24
	(0.15)	(0.13)		(0.10)
coef2	-0.29	-0.33		-0.27
	(0.15)	(0.12)		(0.10)
$\sum(\mathbf{coef1}+\mathbf{coef2})$	-0.05	0.04		-0.03
c			0.18	
$\mathbf{R}^2$	0.16	0.13	0.19	0.09
SSE			0.23	0.27
N	87	77	164	164
p-value for het.test		0.12		0.002

Values in the brackets are the heteroscedasticity corrected standard errors.

5 Testing for the AR model against the TAR alternative

Although our main purpose is to evaluate the models by their forecasting ability, we start by discussing the statistical choice between them in sample.

Conventional tests of the null of an AR model against the TAR alternative have non-standard distributions, since the threshold parameter is not identified under the null. In empirical applications frequently the p-values for the LR statistic (AR null versus the TAR alternative) are generated by Monte Carlo methods (e.g. Coakley-Fuertes, 1998; O'Connell-Wei, 1997; Obstfeld-Taylor, 1997).

In this study, we follow Hansen (1996a), who develops a method to replicate the asymptotic distribution of the test statistic. The procedure is described in Hansen (1996b) as well. Our null hypothesis is $H_0 : \lambda = \beta$, where $\lambda = (\lambda_0\ \lambda_1\ \lambda_2)$ and $\beta = (\beta_0\ \beta_1\ \beta_2)$, and the alternative is $H_1 : \lambda \neq \beta$. When the threshold is known and the errors are iid, the standard F-statistic could be used to test the null against the alternative. The F statistic is

$F_n(c) = n(\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2(c)) \ / \ \widehat{\sigma}_n^2(c)$,

where $\widetilde{\sigma}_n^2$ is the residual variance from estimating the model under the null, ie. fitting an $AR(2)$ model into the data using OLS. $\widehat{\sigma}_n^2(c)$ is the residual variance from estimating the TAR with the threshold fixed at c. The test statistic, $F_n(c)$, is formed for the value of the c that minimizes the $\widehat{\sigma}_n^2(c)$, call it F_n.

In our case, since c is not identified, the asymptotic distribution of F_n is not chi-square. Following Hansen (1996a) the asymptotic distribution will be approximated by a bootstrap procedure as follows:

1. Let u_t^* , $t = 1, ..., n$ be iid $N(0, 1)$ random draws
2. Set Δy_t^* equal to u_t^*
3. Regress Δy_t^* on $y_{t-1}, \cdots, y_{t-p}$ to obtain the residual variance $\widetilde{\sigma}_n^{*2}$
4. Regress Δy_t^* on the TAR model with the threshold set at c. The residual variance from this regression is denoted by $\widehat{\sigma}_n^{*2}(c)$. Choose c so that $\widehat{\sigma}_n^{*2}(c)$ is

minimized.

5. Form $F_n^* = n(\widetilde{\sigma}_n^{*2} - \widehat{\sigma}_n^{*2}(c)) \;/\; \widehat{\sigma}_n^{*2}(c)$

Hansen (1996a) shows that the distribution of F_n^* converges weakly in probability to the null distribution of F_n, so that repeated bootstrap draws from F_n^* may be used to approximate the asymptotic null distribution of F_n. The bootsrap approximation to the asymptotic p-value of the test is formed by counting the percentage of bootstrap samples for which F_n^* exceeds the observed F_n. P-value from this exercise turns out to be 0.008 indicating that the null of AR-model can be rejected in favor of the TAR-model.

6 Forecasting out-of-sample

6.1 Forming the TAR forecasts

Judging the out-of-sample forecasting performance of TAR models is essential, since neither statistically significant rejections of linearity nor in sample fit are a quarantee of the usefulness of the model for forecasting. Significant rejections of linearity often occur while the out-of-sample forecasting performance is no better than for a simple linear model. This can be due to in sample non-linearities, outliers or structural shifts (see e.g. Diebold and Nason, 1990). Forecast performance is typically not reported when TAR models are fitted to the real exchange series. This may be partly because it is not possible to obtain closed-form analytic expressions for the h-step ahead forecasts[3].

One method to generate the h-step ahead forecasts for a TAR model is via Monte Carlo simulation. This is done as follows.

1. The data is divided into the training period (estimation period) consisting of the first T observations, and the validation period (testing period) consisting of the remaining n observations. Total number of observations is $T+n$.

2. The training period data is used to estimate the model in (3) giving us the parameter estimates, $(\widehat{\lambda}_0, ..., \widehat{\lambda}_2, \widehat{\beta}_0, ..., \widehat{\beta}_2, \widehat{c})$ and $\overline{y}$. The residuals from this estimation are collected into two vectors corresponding to the residuals from the outer and the inner regime regressions, denoted by e_T^{outer} and e_T^{inner} respectively.

3. The estimated process is used to simulate the TAR process h steps into the future M times as follows.

$$\widetilde{y}_{T+h,j} = \widehat{\lambda}_0 + \widehat{\lambda}_1 \widetilde{y}_{T+h-1,j} + \widehat{\lambda}_2 \widetilde{y}_{T+h-2,j} + \epsilon_{T+h,j}^{outer}, \qquad \text{if} \qquad \widetilde{y}_{T+h-1,j} - \overline{y} > \widehat{c}$$
$$\widetilde{y}_{T+h,j} = \widehat{\beta}_0 + \widehat{\beta}_1 \widetilde{y}_{T+h-1,j} + \widehat{\beta}_2 \widetilde{y}_{T+h-2,j} + \epsilon_{T+h,j}^{inner}, \qquad \text{if} \qquad \widetilde{y}_{T+h-1,j} - \overline{y} \leq \widehat{c}$$

where $j = 1, ..., M$ and $\widetilde{y}$ denote the forecasts.

To do the simulation into the future, we need to decide how to generate the shocks to the inner and outer regime processes, i.e. $\epsilon_{T+h,j}^{inner}$ and $\epsilon_{T+h,j}^{outer}$. Since we do not know what the real error distribution is, we proceed by picking up the errors

[3] E.g. Tiao and Tsay (1994) and Clements and Smith (1997,1999) discuss the out-of-sample forecasting of TAR models.

randomly and uniformly from the estimated residual distributions $e_T^{outer}$and e_T^{inner}. The benefit of this non-parametric approach is that it is robust to possible violations in error assumptions. Alternatively, we could draw the error sample from $N(0, \sigma^2)$ where σ^2 would be the estimated error variance.

Finally, the average of the forecasts for each horizon h over the M iterations are calculated:

$\widetilde{y}_{T+h} = \sum_{j=1}^{M} \widetilde{y}_{T+h,j}$ for $h = 1, ..., H$

These averages over the M iterations are our MC forecasts of $\widetilde{y}_{T+h}$ for H horizons.

The RMSEs are used to evaluate the distance between the forecasted and the actual values. Rather than evaluating the h-step ahead forecast performance based on the difference between only y_{T+h} and $\widetilde{y}_{T+h}$ the exersice is done as follows. The first training period, T, is extended one period at time and the steps from 1 to 3 are done over again. Finally, the resulting forecasts and RMSEs are computed as

$$RMSE_{T,h}^{TAR} = \sqrt{\frac{1}{n-h+1}\sum_{i=0}^{n-h}(y_{T+h+i} - \widetilde{y}_{T+h+i})^2}$$

where $\widetilde{y}_{T+h+i}$ is the h-step ahead forecast of y_{T+i+h} based on $\{y_t\}_1^{T+i}$. The h-step forecast performance is then evaluated by calculating the RMSEs based on all training periods from T to $T+i$. $RMSE_{T,h}^{TAR}$ denotes this measure of h-step forecast performance indexed by the first training period ending point, y_T.

The out-of-sample forecasting performance of the AR model is evaluated similarly by calculating a comparable RMSE measure ($RMSE_{T,h}^{AR}$) indexed by the first training period end point and the forecast horizon. The formulas for the out-of-sample forecasts for $AR(2)$ model are presented e.g. in Hamilton (1994, p.81).

Since forecasting performance is typically sensitive to the forecast origin, we do the forecasting exercise for four different initial training period lengths to check for sensitivity. In short samples of real data there is the trade off between the length of the training period and the validation period. The longer the training period, the better we can expect to estimate the long run real exchange rate and the parameter values, and thus we can expect to get better forecasts. On the other hand, since the focus is on evaluating the forecasting performance, the validation periods need to be long enough to allow for the long tern purchasing power convergence rather than temporary cycles to drive the results.

6.2 Results

The RMSEs for the out-of-sample forecasts of the TAR and AR models, and their ratios are presented in Table 2. It is immediately clear that the forecast performance is sensitive to the length of the initial training period. Likewise none of the models is always better than the other. Some interesting patterns emerge though. Generally, the forecast accuracy improves the longer the initial training period is. Both models appear to perform the best when the period since 1993 is forecasted. The linear AR model is much more robust to the choice of the training period and the forecast horizon than the nonlinear TAR model. The nonlinear forecasts, particularly at the longer horizons, seem to be really bad in some cases relative to the AR model.

Overall, even if the TAR model is doing better with the longer training periods (or in forecasting only the more recent data), its performance does not appear to be significantly better than that of the simple linear AR(2) model – particularly at longer horizons. At shorter horizons, up to three years, there are periods when TAR outperforms the AR model, but overall if you are to choose one method of forecasting, the simple linear model appears to be more reliable.

How can we explain the failure of the threshold model to outperform the simple AR specification? The results may be simply sensitive to the specification of the TAR model. When we look at the estimation results of the TAR for the full data set in table 1, we can see that the two estimated regimes are in fact quite similar. Particularly, the inner regime is not showing convergence at all, which might give us a good reason to impose a unit root for the inner regime to start with. Given that the TAR model requires more parameters to be estimated than the AR model, this may create large enough uncertainty in the coefficient estimates to lead to quite unreasonable long horizon forecasts in some cases.

TABLE 2

Out-of-sample forecast performance: $RMSE_{T,h}^{TAR}$ vs. $RMSE_{T,h}^{AR}$

	Forecast horizon, h				
	4	**8**	**12**	**16**	**20**
First training period:			**1957Q1-92Q4**		
TAR	0.088	0.102	0.093	0.086	0.093
AR	0.090	0.143	0.134	0.056	0.054
TAR/AR	0.98	0.71	0.69	1.54	1.72
Out-of-sample obs			24		
First training period:			**1957Q4-89Q4**		
TAR	0.082	0.082	0.086	0.090	0.103
AR	0.096	0.119	0.126	0.122	0.138
TAR/AR	0.85	0.69	0.68	0.74	0.75
Out-of-Sample obs			36		
First training period:			**1957Q4-84Q4**		
TAR	0.205	0.607	1.56	3.94	10.00
AR	0.118	0.165	0.173	0.160	0.185
TAR/AR	1.74	3.68	9.02	24.63	54.05
Out-of-sample obs			56		
First training period:			**1957Q1-79Q4**		
TAR	0.199	0.534	1.32	3.27	8.24
AR	0.131	0.192	0.224	0.246	0.267
TAR/AR	1.52	2.78	5.89	13.29	30.86
Out-of-sample obs			76		

The rows denoted by TAR and AR report the RMSEs of the out-of-sample forecasts of the TAR and AR models respectively for forecast horizons from one to five years. The rows denoted by TAR/AR report the ratio of the two models RMSEs.

7 Concluding remarks

We set out to compare the out-of-sample forecasting performance of a linear AR model and a nonlinear TAR model. The idea was to see if the intuitively appealing suggestion of nonlinear adjustment of real exchange rates toward purchasing power parity, due to market frictions, could be utilized in forecasting. Indeed, we found the tendency of the US/DEM real exchange rate to converge faster when the previous real exchange rate was farther away from the estimated purchasing power parity level. However, this tendency did not appear to be strong enough for the TAR specification to lead to consistently better out-of-sample forecasting performance. One reason may be that the data sample is not long enough to reliably estimate the non-linear specification which requires more parameters to be estimated than the more parsimonious linear specification.

The results presented in this paper are very preliminary. The next step we want to pursue is to check the robustness of the results to the assumption on the inner regime process. Specifically, we want to a restrict the inner regime to follow a unit root process. This is reasonable given that the estimated inner regime does not appear mean reverting. Also, we would like to tackle the forecast uncertainty further. Recently, Kitamura (1999) has paid attention to the uncertainty involved in the coefficient estimates and the resulting risk involved in the forecasts. He suggests using bootstrap smoothing to reduce the variance of the forecasts. The intuition is that since the forecast is sensitive to the realization of training period samples, its variance is reduced by first perturbing the training observations using the bootstrap and then averaging over the perturbed series. Following this idea we intend to combine bootstrapping with the MC forecasting hoping to reduce the variations in the TAR forecasts and thus potentially improve their overall accuracy. Also, we want to check for the robustness of the results for the assumptions about the error process. Doing the estimations for other currencies would help in gaining some understanding of the generality of the results as well. If our current results are found to hold with these robustness checks, we would conclude that the theoretically and intuitively attractive idea behind the nonlinear threshold adjustment may not be empirically significant enough to help us in understanding the future behavior of real exchange rates.

References

Clements, M. P. – Smith, J. (1999) **A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models.** Journal of Applied Econometrics 14, 123–141.

Clements, M. P. – Smith, J. (1997) **The Performance of Alternative Forecasting Methods for SETAR Models.** International Journal of Forecasting 13, 463–475.

Coakley, J. – Fuertes, A.M. (1998) **TAR Models of European Real Exchange Rates.** Manuscript.

Diebold, F.X. – Nason, J. A. (1990) **Nonparametric Exchange Rate Prediction?** Journal of International Economics 28, 315–332.

Engel, C. (1996) **Long-run PPP May Not Hold After All.** NBER Working paper 5646.

Hamilton, J.D. (1994) **Time Series Analysis**. Princeton University Press.

Hansen, B. (2000) **Sample Splitting and Threshold Estimation.** Econometrica 68, 575–604.

Hansen, B. (1996a) **Inference When a Nuisance Parameter is not Identified under The Null Hypothesis.** Econometrica 64, 413–430.

Hansen, B. (1996b) **Inference in TAR Models.** Manuscript.

Kitamura, Y. (1999) **Predictive Inference and the Bootstrap.** Manuscript.

Kuo, B.-S. – Mikkola, A. (1999a) **Re-Examining Long-Run Purchasing Power Parity.** Journal of International Money and Finance 18, 251–266.

Kuo, B.-S. – Mikkola, A. (1999b) **How Sure are We About PPP? Panel Evidence with The Null of Stationary Real Exchange Rates.** University of Helsinki, Department of Economics Discussion Papers, No. 451.

Lothian, J. (1998) **Some New Stylized Facts of Floating Exchange Rates.** Journal of International Money and Finance 17, 29–39.

Lothian, J. – Taylor, M.P. (1996) **Real Exchange Rate Behavior: The Recent Float from The Perspective of The Past Two Centuries.** Journal of Political Economy 104, 488–510.

Michael, P. – Nobay, A.R. – Peel, D. A. (1997) **Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation.** Journal of Political Economy 105, 862–879.

Obstfeld, M. – Taylor, A. (1997) **Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited.** NBER Working Paper 6053.

O'Connell, P. (1998a) **Market Frictions and Real Exchange Rates.** Journal of International Money and Finance 17, 71–95.

O'Connell, P. (1998b) **The Overvaluation of Purchasing Power Parity.** Journal of International Economics 44, 1−19.

O'Connell, P. − Wei, S-J. (1997) **"The Bigger They Are, The Harder They Fall": How Price Differences Across U.S. Cities are Arbitraged.** NBER Working paper 6089.

Tiao, G.C. − Tsay, R.S. (1994) **Some Advances in Non-linear and Adaptive Modelling in Time-series. Journal of Forecasting.** 13, 109−131.

BANK OF FINLAND DISCUSSION PAPERS

ISSN 0785-3572, print; ISSN 1456-6184, online

1/2000 Jussi Snellman – Jukka Vesala – David Humphrey **Substitution of Noncash Payment Instruments for Cash in Europe.** 2000. 39 p. ISBN 951-686-647-6, print; ISBN 951-686-648-4, online. (TU)

2/2000 Esa Jokivuolle – Samu Peura **A Model for Estimating Recovery Rates and Collateral Haircuts for Bank Loans.** 2000. 22 p. ISBN 951-686-649-2, print; ISBN 951-686-650-6, online. (RM)

3/2000 Risto Herrala **Markets, Reserves and Lenders of Last Resort as Sources of Bank Liquidity.** 2000. 24 p. ISBN 951-686-653-0, print; ISBN 951-686-654-9, online. (TU)

4/2000 Pekka Hietala – Esa Jokivuolle – Yrjö Koskinen **Informed Trading, Short Sales Constraints and Futures' Pricing.** 2000. 29 p. ISBN 951-686-655-7, print; ISBN 951-686-656-5, online. (RM)

5/2000 Mika Kuismanen **Labour Supply and Income Tax Changes: A Simulation Study for Finland.** 2000. 36 p. ISBN 951-686-657-3, print; ISBN 951-686-658-1, online. (TU)

6/2000 Ralf Pauli **Payments Remain Fundamental for Banks and Central Banks.** 2000. 40 p. ISBN 951-686-659-X, print; ISBN 951-686-660-3, online. (RM)

7/2000 Yuksel Gormez – Forrest Capie **Surveys on Electronic Money.** 2000. 46 p. ISBN 951-686-661-1, print; ISBN 951-686-662-X, online. (TU)

8/2000 Markus Haavio – Heikki Kauppi **Housing Markets, Liquidity Constraints and Labor Mobility.** 2000. 26 p. ISBN 951-686-663-8, print; ISBN 951-686-664-6, online. (TU)

9/2000 Ari Hyytinen – Otto Toivanen **Monitoring and Market Power in Loan Markets.** 2000. 49 p. ISBN 951-686-667-0, print; ISBN 951-686-668-9, online. (RM)

10/2000 Ari Hyytinen – Tuomas Takalo **Enhancing Bank Transparency: A Re-assessment.** 2000. 34 p. ISBN 951-686-669-7, print; ISBN 951-686-670-0, online. (RM)

11/2000 David G Mayes – Matti Virén **Asymmetry and the Problem of Aggregation in the Euro Area.** 2000. 39 p. ISBN 951-686-671-9, print; ISBN 951-686-672-7, online. (TU)

12/2000 Erkki Koskela – Rune Stenbacka **Agency Cost of Debt and Lending Market Competition: A Re-Examination.** 2000. 20 p. ISBN 951-686-673-5, print; ISBN 951-686-674-3, online (TU)

13/2000 Biing-Shen Kuo – Anne Mikkola **Forecasting the Real US/DEM Exchange Rate: TAR vs. AR.** 2000. 18 p. ISBN 951-686-675-1, print; ISBN 951-686-676-X, online (TU)