University of Essex

SCHOOL OF ACCOUNTING, FINANCE AND MANAGEMENT

Essex Finance Centre

Fiscal Readjustments in the US: A Non-linear Time Series Analysis

Andrea Cipollini
University of Essex, School of Accounting, Finance and Management

Bassam Fattouh
CeFiMS, SOAS, University of London

Kostas Mouratidis
University of Swansea, Business School

Discussion Paper No. DP 07/04

November, 2007

¥

Fiscal Readjustments in the US: A Non-linear Time Series Analysis

Andrea Cipollini

University of Essex, School of Accounting, Finance and Management

Bassam Fattouh ♣

CeFiMS, SOAS, University of London

Kostas Mouratidis

University of Swansea, Business School

March 2007

Abstract

We analyze the fiscal adjustment process in the US using a multivariate threshold Vector Error Regression Model (VECM). We find that the shift from a single equation to multivariate setting adds value both in terms of our economic understanding of the fiscal adjustment process in the US and the forecasting performance of non-linear models. First, we find evidence that fiscal authorities will intervene to reduce real per capita deficit only when it reaches a certain threshold and that the fiscal adjustment process takes place primarily by cutting government expenditure rather than increasing tax revenues. Second, the out-of-sample density forecast and probability forecasts results suggest that a shift from a univariate AR model specification to a multivariate model improves forecast performance. We also find that the forecasting performance of both linear and non-linear VECM is similar for long horizons (e.g. two years ahead).

Keywords: Threshold Cointegration, Forecasting, Deficit Sustainability

JEL classification: C32, C53, E62

Acknowledgments: The authors which to thank three anonymous referees. All the computations have been carried using GAUSS. The authors wish to thank Serena Ng and Hansen and Seo for making available the GAUSS routines used to for unit root test and for the estimation of a threshold VECM, respectively.

♣Corresponding Author: Department for Financial and Management Studies, SOAS, University of London, Thornhaugh Street, Russell Square, London WC1H 0XG. Tel. 0044-(20)78984053; Fax 0044-(20) 78984089. e-mail: bf11@soas.ac.uk

1. Introduction

The recent deterioration in US budget deficits has raised serious concerns about the long-run sustainability of the U.S. fiscal policy. In addressing this issue, many studies have examined whether the U.S. fiscal policy respects the inter-temporal government budget constraint. This constraint implies that Ponzi games in which the government rolls over its debt in full every period by borrowing to cover both principal and interest payments are ruled out as viable option for government finances. The no-Ponzi game restriction, which is regarded as synonymous with sustainability, requires that today's government debt is matched by an excess of future primary surpluses over primary deficits in present value terms. This condition imposes testable restrictions on the time series properties of key fiscal measures such as the stock of public debt, the budget deficit and the long run relationship between government expenditures and revenues. In a seminal article, Hamilton and Flavin (1986) suggest that a sufficient condition for the inter-temporal budget constraint to hold is for the deficit inclusive of interest payments to be stationary. Wilcox (1989) extends the work of Hamilton and Flavin by allowing stochastic interest rates and non-stationarity in the non-interest surplus. He shows that when the sustainability condition holds, the present value of the stock of public debt should be stationary and has an unconditional mean of zero. Trehan and Walsh (1988) generalize the Hamilton and Flavin result and show that if debt and deficits are integrated of order one and if interest rates are constant, then a necessary and sufficient condition for sustainability is that debt and primary balances (net-of-interest deficits) are cointegrated. Other studies examine the time series properties of government spending and revenues. For instance, Hakkio and Rush (1991) show that a necessary condition for inter-temporal budget constraint is the existence of cointegration between government expenditure (inclusive of interest payments) and government revenues. Quintos (1995) expands on Hakkio and Rush (1991) and introduces the concept of strong sustainability condition which implies that the undiscounted public debt is finite in the long run.

More recent work has emphasized the importance of non-linearity in the US fiscal policy. This non-linearity may arise if we expect fiscal authorities to react differently to whether the deficit has reached a certain threshold deemed to be unacceptable or unsustainable. Bertola and Drazen (1993) develop a framework which allows for trigger points in the process of fiscal adjustment, such that significant adjustments in budget deficits may take place only when the ratio of deficit to output reaches a certain threshold. This may reflect the existence of political constraints that block deficit cuts, which are relaxed only when the budget deficit reaches a sufficiently high level deemed to be unsustainable (Bertola and Drazen, 1993; Alesina and Drazen, 1991). Recent studies have found strong evidence of non-linearity in US fiscal policy. Using an Exponential Smooth Transition Auto-regressive model (ESTAR) and long span data set starting from 1916, Sarno (2001) provides evidence of non-linear mean reversion in the US debt-GDP ratio. By using a Threshold Autoregressive (TAR) model, Arestis et al (2004) provide evidence of threshold effects in the US per capita deficit such that policymakers will intervene to reduce per capita deficit only when it reaches a certain threshold.

In line with the above studies, we provide new evidence of strong non-linearity in the US fiscal policy. We contribute to the existing literature by extending the analysis of US fiscal adjustment from a single equation setting to a multivariate one, using a non-linear Vector Error Correction model. This extension adds value both in terms of our economic understanding of the fiscal adjustment process in the US and the forecasting power of the model. First, using a multivariate threshold cointegration model, we are able to identify whether the government's solvency constraint in the US is achieved through revenue increases, spending cuts or a combination of both. The issue of which specific item of the budget ensures fiscal re-adjustments has received considerable attention among US policymakers and has been recently the focus of much heated debate. For instance, Rubin, Orszag and Sinai (2004) argue that "balancing the budget for the longer term will require a combination of expenditure restraint and revenue increases". The authors believe that "the single most important act Congress and the Administration could take at this point to rein the budget over the next decade would be to re-establish the budget rules that existed in the 1990s. These put caps on discretionary spending and required that reductions in taxes or increases in mandatory spending be paid for with other tax increases or spending cuts". A study by the Congressional Budget Office (2003) also cautioned that "economic growth alone is unlikely to bring the nation's long term fiscal position into balance".

The contribution of the academic literature to this debate has been very limited. Alesina and Perotti (1995) find evidence that for fiscal adjustment to be permanent and effective, the focus must be on level of expenditure rather than taxation.[1] They argue that tax increases ease fiscal problems only temporarily. Temporary tax increases may also be very difficult to reverse and as such tax-driven deficit cuts may induce high tax ratios. Furthermore, raising taxes is unpopular and there are doubts whether such a strategy can in fact increase government revenues. Bohn (1991) and Crowder (1997) rely on the government inter-temporal solvency condition to analyze the performance of fiscal stabilization plans over a long-term data span. Specifically, the budget item series showing most of the error correcting dynamics is the one baring most of the fiscal re-adjustment burden**.** Crowder (1997) shows that the large U.S. deficits in the 1980s and early 1990s have been primarily caused by increases in government spending rather than falls in tax revenues. Thus, in order to restore the inter-temporal budget constraint, the bulk of fiscal readjustment should occur through government spending cuts rather than through increases in tax revenues. Bohn (1991) shows that regardless of the shock that caused the high budget deficit, historically these deficits have been corrected by combination of both

[1] Alesina and Perotti (1995) use the long-run, cyclically adjusted primary deficit to identify periods of fiscal re-adjustment. Specifically, a very tight fiscal policy in year t occurs when the cyclically adjusted deficit decreases by more than 1.5 percent of GDP. A successful fiscal adjustment in year *t* occurs when a tight fiscal policy implemented in year is such that the gross debt-to-GDP ratio in year t + 3 is at least 5 percentage points lower than in year t.

spending cuts and tax increases. Auerbach (2000) finds that both components of U.S. fiscal policy have been responsive to the fluctuations in the deficit although the response from government spending has been much more important.

Our results reveal the following important findings. They provide support for the existence of trigger points in U.S. fiscal policy. Specifically, we find strong evidence of non-linearity in the fiscal process where adjustment occurs only when the real deficit per capita reaches a certain threshold. Below this threshold, there seem to be no significant error-correction effects, which may suggest that policymakers become sensitive to large deficits only when the deficit reaches very 'high' level deemed to be unacceptable or unsustainable. More importantly, we find that government expenditure shows the strongest error correcting dynamics and hence the bulk of fiscal adjustment seems to occur through spending cuts rather than increases in tax revenue.

In addition to gaining better understanding of the US fiscal adjustment process, we also evaluate the out-of-sample density forecast and probability forecast performance of the estimated model. Our results highlight an additional advantage from generalizing the model from a single equation to multivariate setting. Specifically, the results from out-of-sample density forecast and probability forecasts suggest that there is an improvement in the model forecast performance once we move from a univariate AR model specification to a multivariate model. We also compare the out-of-sample forecast performance of the linear and threshold model. In a recent survey, Granger (2001) concludes that a major weakness of the literature on non-linear models is that little is known about the out-of-sample forecasting properties of different non-linear models or their out-of-sample forecast performance with those corresponding to linear models. The empirical findings suggest that, although the threshold VECM has a slight better probability forecast performance than the linear VECM, the density forecast performance of both the linear and non-linear VECM is similar for the long horizon (e.g. two years ahead) and thus we can not recommend the use of the threshold VECM over simple linear models for forecasting purposes. Similar results have been found recently in the context of the exchange market (see for instance, Rapach and Wohar, 2006). This suggests that although non-linear models are useful to gain better understanding of the US fiscal policy, they do not necessarily provide more reliable forecasts.

This paper is organized as follows. Section 2 describes the empirical methodology while section 3 presents the empirical results. Section 4 summarizes and concludes.

2. Empirical Method

2.1 Threshold Cointegration

A Vector Error Correction model, VECM, fitted to both *G*, the real government expenditure per capita and to *R,* the real government revenue per capita, can be used to test whether there is any evidence of public finance sustainability and to test which of the two fiscal series carries the burden of fiscal readjustment (if any). Many empirical studies have concentrated on estimating the following linear VECM (where, for simplicity, we fix to one the VECM lag order):

$$\begin{pmatrix} \Delta G_t \\ \Delta R_t \end{pmatrix} = \mu + \alpha w_{t-1} + \Gamma \begin{pmatrix} \Delta G_{t-1} \\ \Delta R_{t-1} \end{pmatrix} + u_t \qquad (1)$$

with μ a two dimensional vector of intercepts, $w_{t-1}=G_{t-1}-\beta R_{t-1}$, α is a two dimensional vector of speed of adjustment coefficients, and u_t is the error term vector. According to Quintos (1995), the deficit is 'strongly' sustainable if the I(1) processes R_t and G_t are cointegrated and $\beta = 1$, while it is 'weakly' sustainable if R_t and G_t are cointegrated and $0 < \beta < 1$. Weak sustainability implies that the government constraints holds, but the undiscounted debt process is exploding at a rate that is less than the growth rate of the economy. Although this case is consistent with sustainability, it is inconsistent with the ability of the government to market its debt in the long run. Thus, in this paper, we will only test for the 'strong' sustainability condition and set $\beta =1$.[2] By setting $\beta =1$, the error correction term becomes the real deficit per capita.

As argued above, equation (1) may not be the most appropriate means to characterize the fiscal adjustment process for there may exist trigger points in the process of fiscal adjustment. Hence, in this study, we focus on the following threshold VECM:

$$\begin{aligned} \begin{pmatrix} \Delta G_t \\ \Delta R_t \end{pmatrix} &= \mu_1 + \alpha_1 w_{t-1} + \Gamma_1 \begin{pmatrix} \Delta G_{t-1} \\ \Delta R_{t-1} \end{pmatrix} + u_{1t}, w_{t-1} \le \gamma \\ \begin{pmatrix} \Delta G_t \\ \Delta R_t \end{pmatrix} &= \mu_2 + \alpha_2 w_{t-1} + \Gamma_2 \begin{pmatrix} \Delta G_{t-1} \\ \Delta R_{t-1} \end{pmatrix} + u_{2t}, w_{t-1} > \gamma \end{aligned} \qquad (2)$$

The model given (2) allows us to test whether there are significant asymmetries in the adjustment process of per capita government revenues and per capita government expenditure to the long-run equilibrium level depending on the level of deficit per capita, w_{t-1}, given by G_t - R_t. In particular, if the real deficit per capita exceeds the trigger point γ, then there is a switch in the speed of adjustment coefficients from α_1 to α_2, as well for the other short-run dynamics parameters. Hansen and Seo (2002) suggest to estimate the model given by (2) through Maximum Likelihood under the assumption that the errors u_t are iid Gaussian. The Gaussian likelihood is:

$$L_n = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{n} u_t{}'\Sigma^{-1}u_t \qquad (3)$$

where:

[2] Most recent empirical studies also suggest evidence of strong sustainability either without regime shifts (see Cunado et al, 2004) or with regime shifts (see Martin, 2000, and Arestis et al, 2004).

$$u_t = \begin{pmatrix} \Delta G_t \\ \Delta R_t \end{pmatrix} - \left(\mu_1 - \alpha_1 w_{t-1} - \Gamma_1 \begin{pmatrix} \Delta G_{t-1} \\ \Delta R_{t-1} \end{pmatrix} \right) d_{1t}(\gamma) - \left(\mu_2 - \alpha_2 w_{t-1} - \Gamma_2 \begin{pmatrix} \Delta G_{t-1} \\ \Delta R_{t-1} \end{pmatrix} \right) d_{2t}(\gamma)$$

with the indicator function $d_{1t}(\gamma)$ taking value 1 if the deficit is below the trigger point γ, and zero otherwise. Furthermore, $d_{2t}(\gamma)$ is equal to (1 - $d_{1t}(\gamma)$). In order to detect non-linearity, Hansen and Seo (2002) use an LM statistics to test H_0 (linear cointegration) versus H_1 (threshold cointegration). If the cointegrating vector is known and equal to β_0 (in our study is fixed to unity), then the LM test is given by:

$$SupLM^0 = \sup_{\gamma_L \leq \gamma \leq \gamma_U} LM(\beta_0, \gamma) \tag{4}$$

Given that the asymptotic critical values of the distribution of the test statistics cannot in general be tabulated, bootstrapped p-values are computed using both a fixed regressor and a parametric bootstrap (for a description, see Hansen and Seo, 2002).

2.2 Out-of-Sample Density Forecasts

To further motivate the use of threshold VECM, we explore whether our proposed model is superior both to the univariate model and the linear model in terms of its out-of-sample forecast performance. Traditionally, evaluating the forecast accuracy of models has been based on point forecasts using often the Root Mean Square Error (RMSE). The empirical evidence often suggests that the forecasting ability of linear models outperforms non-linear model on the basis of RMSE criterion alone[3]. Several studies however have recently emphasized the importance of evaluating forecast performance on the basis of an estimate of the complete probability distribution of the possible future outcomes of the series (that is a density forecast) as opposed to point forecasting. More specifically, only under certainty equivalence (e.g. policymakers with quadratic loss function and a linear dynamics of predicted variable), the RMSE can be used as a criterion to choose an optimal forecast[4]. If certainty equivalence does not hold, then it is important to focus not only on the first moments, but on the overall density of forecasts. The density forecasts are generated through stochastic simulation and we give in the Appendix a detailed description of this method. First, we produce the density forecasts for both changes in government spending and tax revenues, using a univariate AR model. Then, we produce the marginal density forecasts for both changes in government spending and tax revenues, ΔG and ΔR, respectively. We also produce the conditional density of government spending changes and of tax revenues changes, $\Delta G/\Delta R$ and $\Delta R/\Delta G$, respectively. Finally, we produce the joint density of

[3] Diebold and Nason (1990) give four reasons for why, although non-linear models have better in-sample fit than linear models, they may fail to dominate in terms of out-of-sample forecast performance based on the RMSE (see also Clements and Smith 2000).

[4] Chistofffersen and Diebold, (1997) show that under asymmetric loss the optimal forecast is the conditional mean plus a bias term which depend both on forecaster's loss function and on the conditional variance of predicted variable.

government spending changes and of tax revenues changes, *(ΔG/ΔR)*ΔR* and *(ΔR/ΔG)*ΔG*, respectively. We consider three different forecast horizon, *h*, equal to 1, 4, and 8 quarters ahead, respectively. For the purpose of density forecast evaluation, in line with Clements and Smith (2000), for a given forecast horizon *h*, we calculate the probability integral transforms of the actual realizations, y_t, of each fiscal series over the forecast evaluation period with respect to the model's forecast densities, given by $\{p_t(y_t)\}_{t=1}^{n}$. Therefore, we evaluate the probability integral transform, *PIT*:

$$PIT = \int_{-\infty}^{y_t} p_t(u)du \tag{5}$$

for $t = 1,\ldots,n$. When the model forecast density corresponds to the true predictive density, the sequence z_t is iid, *U(0,1)*. In line with Diebold et al. (1998) and with Clements and Smith (2000), we use informal data analysis to test whether *PIT* is iid, *U(0,1)*. Therefore, the evaluation of accuracy of density predictions consists of assessing uniformity using PP plots[5]. Specifically, we plot the empirical distribution function of *PIT* against the 45^0 line, with critical values defining the confidence intervals obtained from Miller (1956). Then, in order to assess whether the *PIT* series are iid, we use the Langrage Multiplier, test for the null of serial independence of $(PIT - \overline{PIT})^j$ for integer *j* up to order 3, with $\overline{PIT}$ being the mean the probability integral transform series[6].

Furthermore, we consider the Berkowitz (2001) approach to evaluate the accuracy of density forecasts[7]. Specifically we take the inverse of the Gaussian cumulative distribution function with respect to each component of the sequence *PIT* which gives PIT^*. Under the null of iid *U(0,1)* for the sequence *PIT*, the series PIT^* becomes a standard Gaussian random variable. In order to test for normality in PIT^*, Berkowitz (2001) suggested a likelihood ratio test for the joint null of normality and iid in PIT^*. The test statistic is $LR_B = -2[L(0,1,0) - L(\hat{c},\hat{\sigma},\hat{\rho})]$, where $L(\hat{c},\hat{\sigma},\hat{\rho})$ is the value of the maximum likelihood function of an AR(1) model fitted to PIT^*, where $\hat{c}$ and $\hat{\rho}$ are the estimated intercept and autoregressive coefficient and $\hat{\sigma}$ is the estimated standard deviation for the residuals of the AR(1). Under the null, the LR_B has a χ_3^2 distribution.

---

[5] PP plots provide a visual inspection of the discrepancy between shapes created by the patterns of points on a plot and a reference straight line.
[6] A high order is chosen because as noted by Diebold et al (1998) dependence may be present in higher moments.
[7] Recently, an alternative approach to evaluate the accuracy of density forecast has been suggested by Sarno and Valente (2004).

The stochastic simulation is not only used to produce forecasts under any type of scenario (e.g. the density forecast), but also to generate the forecasts for particular type of scenarios. Specifically, we are interested in generating the probability forecasts for two types of events (see Clements, 2005, and Galvao, 2006, for probability forecast analysis). The first one is defined by negative changes in government spending, and the second one is defined by positive changes in tax revenues. Using the simulation method described in the appendix, we produce 1000 *h* step ahead forecast for government spending changes (conditional on the available information set) and we count how many of these forecast are negative. This number divided by 1000 gives the probability forecast for the government spending series. The same methodology is applied to generate the probability forecast for the tax revenue series. We repeat this exercise by increasing the overall sample by one additional observation, till we reach the end of the forecast evaluation period. We use the following indicators of probability forecast accuracy (see Galvao, 2006):

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - R_t)^2$$

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}[(1-R_t)\ln(1-P_t) + R_t \ln(P_t)]$$

where P_t and R_t are the probability forecast and the actual realisation of the variable one is interested in predicting. Finally, the *QPS* score ranges from 0 to 2, with 0 being perfect accuracy. The second one ranges from 0 to ∞. *LPS* and *QPS* imply different loss functions with large mistakes more heavily penalized under LPS.

3. Empirical Analysis

3.1 Data and Data Sources

The dataset used in this study comprises quarterly observations over the period 1947:2 to 2004:4. We examine the dynamics of real per capita expenditure and real per capita revenues and hence we only focus on the strong sustainability condition (see Quintos (1995) for details). We first collect data on the nominal current federal expenditure (inclusive of interest payments) and current federal revenues (seasonally adjusted). We deflate both series by implicit GDP deflator to obtain real values. The series are then deflated by population to obtain real per capita government expenditure and real per capita government revenues. All the data have been obtained from the FRED database available from the Federal Reserve Bank of St. Louis.

3.2 In-Sample Forecasting Analysis

The augmented Dickey-Fuller (ADF) and the Philips-Perron (PP) tests for the null of unit root (see the first two columns of Table 1) suggest that we can not reject the null hypothesis of non-stationarity in the levels of real per capita government expenditure and real per capita government revenue. These findings are also confirmed by the tests developed by Ng and Perron (2001) under *GLS* detrending, using the modified AIC information criterion to select the optimal lag order. Specifically, as for the tax revenue series, the $MZ^{GLS}{}_{a}$, and ADF^{GLS} tests suggest that we cannot reject the null of unit root at any significance level. As for the tax revenue series, according to the $MZ^{GLS}{}_{a}$, we cannot reject the null of unit root, whereas, using the ADF^{GLS} , we cannot reject the null of unit root at 1% significance level.

Before carrying cointegration analysis, we select the VECM lag length. The results are reported in Table 2a for the linear VECM and Table 2b for the threshold VECM. As can be seen from these tables, both the AIC and BIC statistics pick a lag of one. This holds both for the linear and the threshold VECM.[8]

We next test for existence of threshold effects in the VECM using the *SupLM* statistic. As can be seen from Table 3, the $SupLM^{0}$ statistic suggests a strong presence of threshold effects where the null hypothesis of no threshold can be rejected at the 5% level. The Wald tests also point in the same direction. The null hypothesis that the error-correction coefficients and dynamic coefficients are the same in both regimes can be rejected at 5% and 1% levels respectively.

The parameter estimates were calculated by minimization of $\log\left|\hat{\Sigma}(\gamma)\right|$ over a 300 grid points for the parameter γ. The estimates are reported in Table 4. The estimated threshold is 8.859 dollar per capita which implies that the first regime occurs when the real deficit per capita is less or equal than \$8.859. This regime contains 82% of the sample observations. The second regime occurs when the real deficit per capita is above the threshold of \$8.859. Following Hansen and Seo (2002), we label the first regime as the "typical" regime and the second regime as the "unusual" regime. The results in Table 4 show that the typical regime has no significant error correction effects with the coefficients on the lagged error correction terms in both equations ΔR_t and ΔG_t are insignificant at the conventional levels. In contrast, error correction effects occur only in the extreme regime i.e. when the real deficit per capita has risen above the estimated threshold. Interestingly, the results indicate that fiscal re-adjustment occur through spending cuts rather than increases in government revenue: while the estimated coefficient on the error-correction term in the government expenditure equation is large and highly significant, the estimated coefficient on the error correction term in the revenue per capita equation is quite small and not significant at the conventional levels.

---

[8] For robustness, we also estimated the VECM with 2 lags. The results are very similar to those obtained with one lag and to save space we do not report them. The results are available from the authors upon request.

In Figure 1, we plot the deviations of the real deficit per capita form the estimated threshold point estimate over the sample period. Note that in this figure, positive values identify the "unusual" regime, whereas the negative values identify the "typical" regime. Figure 1 clearly shows that there have been four major shifts from the 'typical' to the 'unusual' regime in the real deficit per capita dynamics. First, a major shift occurred in 1975:2, which is the peak of the 1973 oil crisis, which plunged the US economy into a deep recession. A second major shift occurred in 1981:3. This shift, which occurred during Reagan presidency, corresponds to the effects of the legislation passed by the Congress aimed at cutting personal income taxes over the next three years (the 1981 Economic Recovery Tax Act). Since the tax cuts were not met by equal cuts in government spending, the federal budget went into large deficits and remained so for a considerable period of time. It is only in 1987:3 that we witness a regime shift back towards the typical regime. This switch reflects in part the intensive political and economic debate in the Congress and in the media and the efforts made by fiscal authorities to reduce the large and growing budget deficit. These efforts were manifested in the Tax Reform Act of 1986 and the Balanced Budget and Emergency Deficit Control Act, which called for progressive reduction in the deficit and the achievement of a balanced budget by the early 1990s (Ippolito, 1990). Despite the efforts made to balance the budget, another major shift (the third one) from the 'typical' to the 'extreme' regime occurred in 1991:2. This switch occurred during the Senior Bush presidency and corresponds closely to the recession that plunged the U.S. economy at the beginning of Senior Bush's term and later to the budgetary requirements of the Gulf War. In 1994:2, there was a switch to the typical regime, which lasted for the rest of the 1990s. This coincided with President Clinton's move to the White House and the importance he has attached to balancing the budget in his economic policy. Finally, in 2002:4, there was a switch from the typical to the unusual regime. This switch corresponds to current President's Bush presidency with its emphasis on cutting taxes and boosting defense and security outlays which caused large budget deficits.

3.3 Out-of-Sample Forecasting Analysis

We compare the out-of-sample forecast performance of the linear model and the threshold cointegration model. We leave out the last 64 observations of the sample for density forecast evaluation. More specifically, the forecast evaluation period starts from 1989:1 which corresponds to the beginning of the George Bush senior administration and ends in 2004:4.

In order to produce out-of-sample forecasts, we estimate recursively the three different model specifications (univariate AR, linear and non-linear VECM). We concentrate on one quarter, one year and two years ahead predictions. As for the one quarter ahead projections, we consider, initially, the sample that ends in 1988:4, and then we increase the sample by one observation each time period till we reach a sample period that ends in 2004:3. In order to produce four quarters ahead predictions, we consider, initially, the sample that ends in 1987:4, then we increase the sample by one observation each time period till we reach a sample period that ends in 2003:4. Finally, to produce eight quarters ahead predictions, we consider, initially, the sample that ends in 1986:4, then we increase the sample by one observation each time period till we reach a sample period that ends in 2002:4.

The out-of-sample point forecast evaluation in Table 5 shows that the evidence is inconclusive: the Root Mean Square Error (RMSE) corresponding to the point forecast of the government spending series obtained from the different models are close to each other at the different forecast horizons. Moreover, although the non-linear VECM is the worst in the one quarter ahead point prediction of tax revenues, the different models have a similar performance for the one and two year forecast horizon.

The results from Table 6a-6c suggest that none of the models proposed is capable of providing a good density forecast for the tax revenues series. Specifically, although the PP plots for the probability integral transform sequence (see the right hand side panel of Figures 2-8), show the 45^0 line inside the confidence interval bands for all the model specifications and for most of the forecast horizon (with the exception of the one year ahead density forecast from the univariate AR model, see Figure 2), the Lyung-Box test suggests evidence of serial correlation in the first and third moments of the *PIT* sequence.[9] As for government spending, there is an improvement in density prediction performance once we move from the univariate AR model to the multivariate model and as we consider a forecast horizon longer than one quarter. In particular, even though there is no evidence of serial correlation in the first, second and third moment of the *PIT* sequence corresponding to AR density forecasts (see Table 6a-6c), the corresponding PP plots show the 45^0 line outside the confidence interval bands (see the left hand side panel of Figures 2), for density prediction over one and two years respectively.

As for the multivariate models, the density forecast performance for the linear and for the threshold VECM specification is similar for long horizon (e.g. two years ahead). The Ljung Box test suggest absence of serial correlation in the first, second and third moment of the *PIT* sequence for the marginal, conditional and joint density forecast of government spending produced by both the linear and non-linear VECM and for any forecast horizon (see Table 6a-6c). However, the PP plots for the probability integral transform associated with the threshold VECM marginal, conditional and joint density forecast of government spending have the 45^0 line inside the confidence interval bands only when we consider an eight step ahead forecast horizon (see Figures 6,7 and 8). From Figures 3, 4 and 5, we can observe that the PP plots for the probability integral transform associated with the linear VECM marginal, conditional and joint density forecast of government spending have the 45^0 line inside the confidence interval bands for any forecast horizon.

Using the Berkowitz (1999) test, from table 7a-7c we can observe that the strongest rejection of the null hypothesis of normality and iid for the inverse of the cumulative (standardised) Gaussian

[9] It is worth noting that empirical distribution type of tests, such as PP plots, are valid only under the assumption that *PIT* follows an i.i.d. process (see Spanos, 1999).

distribution with respect to the PIT sequence applies to the AR and Linear VECM. When we use a Threshold VECM, there is a mild non rejection of the null hypothesis if we turn our focus on the conditional and joint density (one step ahead) forecast of government spending changes and on the marginal and joint density (eight step ahead) forecast of government spending changes.

Finally, the probability forecast exercise confirms the results obtained from the density forecast evaluation. As mentioned in section 2.2, we are interested in evaluating the model forecast performance regarding events which can be associated with fiscal readjustments and these are either positive changes in tax revenues or negative changes in government spending. Therefore, as mentioned in section 2.2, we need to compute probability forecasts and evaluate them in terms of QPS and LPS scores. As for government spending (see Table 8a), the best performer (in terms of QPS and LPS scores) for any type of prediction horizon is the non-linear VECM model are considerably lower than the corresponding one for the AR model. As for the tax revenues (see Table 8b), the worst probability forecast performance is the one associated with the non-linear VECM for the one quarter ahead probability forecast. There are gains from moving to a univariate AR modelling framework to a multivariate model if the prediction horizon is either one or two years ahead, and the non-linear VECM is the best performer (in terms of QPS and LPS scores) if the forecast horizon is two year ahead.

4. Conclusions

In this paper, we investigate empirically the US government inter-temporal solvency condition and assess whether government solvency constraint has been achieved mainly through revenue increases or spending cuts or a combination of both. Using a Threshold Vector Error Correction estimation procedure, we find evidence that government authorities would intervene only when the deficit per capita has reached a certain threshold. Our results show that the bulk of fiscal adjustment occurs through spending cuts rather than increases in tax revenue.

In terms of forecasting, the picture is mixed. By evaluating the out-of-sample density forecast performance of the estimated model, we show that there is an improvement in the model forecast performance once we move from a univariate AR model specification to a multivariate model. However, we find that the forecasting performance of both linear and non-linear VECM is similar for long horizon (e.g. two years ahead) and thus we can not recommend the use of the threshold VECM over simple linear models for forecasting purposes. This suggests that our proposed model could be improved upon and should be evaluated in comparison not only with alternative multivariate non-linear models but also multivariate linear models with structural breaks. One might also consider a time trend or an indicator of the US business cycle as an additional threshold variable (beyond the government deficit) in the non-linear multivariate model. Recently, Galvao (2006) have found a good forecasting performance of the US terms spread regarding the US industrial production, using a threshold VAR using both a time trend and the term spread as threshold variables. These extensions can prove very fruitful avenues for future research.

References

Alesina, A. and Drazen, R. (1991), “Why are stabilisation delayed?", *American Economic Review*, Vol. 82, pp. 1170-1188.

Alesina, A. and Perotti, R. (1995), “Fiscal Expansion and Adjustment in OECD countries", *Economic Policy*, pp. 206-247.

Arestis, P., A. Cipollini and B. Fattouh (2004), “Threshold Effects in the U.S. Budget Deficit”, *Economic Inquiry*, Vol. 42, No.2, 214-222.

Auerbach . A J. (2000), “Formation of fiscal policy: the experience of the past twenty-five years” *FRBNY Economic Policy Review*, April, pages 9-23

Balke, N.S. and T.B. Fomby (1997), “Threshold Cointegration”, *International Economic Review,* 38, 627-645.

Berkowitz, J. (2001) “Testing density forecasts, with applications to risk management”, *Journal of Business and Economic Statistics*, 19, 465-474.

Bertola, G. and A. Drazen (1993), “Trigger Points and Budget Cuts: Explaining the Effects of Fiscal Austerity”, *American Economic Review*, 83 (1), 11-26.

Bohn, H. (1991): “Budget balance through revenue or spending adjustments*?”, Journal of Monetary Economics*, 27, pp.333-359.

Chistofffersen, P.F., and Diebold, F., (1997), “Optimal Prediction under Asymmetric Loss”, *Econometric Theory*, Vol. 13, pp. 808-817.

Clements, M. P. (2005): Evaluating Econometric Forecasts of Economic and Financial Variables. Palgrave McMillian eds.

A. B. Galvao (2006): “Structural Break Threshold VARs for predicting the probability of US recessions using the spread” (2006) *Journal of Applied Econometrics.* 21: 463-487.

Clements, M. P. and J. Smith, (2000) “Evaluating the forecast densities of linear and non linear models applications to output growth and unemployment”, *Journal of Forecasting*, 19, pp. 255-276.

Congressional Budget Office (2003), “The Long-Term Budget Outlook.” December.

Crowder, W. (1997), “The U.S. federal intertemporal budget constraint : restoring equilibrium trough increased revenues or decreased spending”, manuscript.

Cunado, J., L. A. Gil-Alana and F. Pérez de Gracia (2004), Is the US fiscal deficit sustainable? A fractionally integrated approach, *Journal of Economics and Business*, 56, No.6, 501-526.

Dickey D.A. and Fuller W.A. (1979): “Distribution of the estimates for autoregressive time series with a unit root”, *Journal of the American Statistical Association*, 74, pp. 427-431

Diebold, F. X., T.A. Gunther, and A. S. Tay, (1998), “Evaluating Density Forecast”, *International Economic Review*, Vol. 39, 863-883.

Diebold, F.X. and J.A. Nason (1990), “Non Parametric Exchange Rate Prediction”, *Journal of International Economics*, 28, 315-332.

Enders and Granger (1998): “Unit Root Test and Asymmetric Adjustment with an Example Using the Term Structure of the Interest Rates”, *Journal of Business and Economics Statistics*, 16, pp. 304-311.

Granger, C. W. J. (2001), An overview of nonlinear macroeconometric empirical models, *Macroeconomic Dynamics*, Vol. 5, No. 4, pp. 466-81.

Hakkio, C.S. and M. Rush (1991), “Is the budget deficit “too large”?”, *Economic Inquiry*, 29, 429-445.

Hamilton, J. D. and M.A. Flavin (1986), “On the Limitation of government borrowing: a framework for empirical testing”, *The American Economic Review*, 76, 808-819.

Hansen, B. and Seo B. (2002): "Testing for two regime threshold cointegration in vector error correction models", *Journal of Econometrics*, 110, pp.293-318.

Haug, A. (1995), "Has Federal budget deficit policy changed in recent years", *Economic Inquiry*, Vol.33, pp.104-118.

Ippolito, D.S. (1990), Uncertain legacies: federal budget policy from Roosevelt through Reagan. Charlottesville : University Press of Virginia.

Martin, G.M. (2000), "US deficit sustainability: a new approach based on multiple endogenous breaks", *Journal of Applied Econometrics*, 15, 83-105.

Miller, L. H. (1956): "Table of percentage points of Kolmogorov statistics", *Journal of American Statistical Association*, 51, pp. 111-121.

Ng, S. and Perron, P. (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica*, 69(6), 1519-1554.

Philips P. and Perron P. (1988): "Testing for a unit root in time series regression", *Biometrika*, 75, pp.335-346

Quintos, C.E. (1995), "Sustainability of the deficit process with structural shifts", *Journal of Business and Economic Statistics*, 13, 409-417.

Rapach, D. and M. Woher (2006) "The Out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior", *International Journal of Forecasting*, Vol. 22, No. 2, pp. 341-361.

Rubin, R.E., P. R. Orszag, and A. Sinai (2004) "Sustained Budget Deficits: Longer-Run U.S. Economic Performance and the Risk of Financial and Fiscal Disarray", Paper presented at the AEA-NAEFA Joint Session, Allied Social Science Associations Annual Meetings.

Sarno, L. (2001), "The behavior of US public debt: a non-linear perspective", *Economics Letters*, 74, 119-125.

Sarno, L. and Valente G. (2004) "Comparing the accuracy of density forecasts from competing models", *Journal of Forecasting*, vol. 23(8), 541-557.

Spanos, A., (1999), Probability Theory and Statistical Inference: Econometric Modeling with Observational Data, Cambridge University Press.

Stephens, M.A. (1974), "EDF Statistics for Goodness-of-fit and Some Comparisons", *Journal of American Statistical Association*, Vol. 69, pp. 730-737.

Tanner, E. and P. Liu (1994), "Is the budget deficit "too large": some further evidence", *Economic Inquiry*, 32, 511-518.

Trehan, B., and C.E. Walsh (1988), "Common Trends, the government budget constraint and revenue smoothing", *Journal of Economic Dynamics and Control*, 12, 425-444.

Trehan, B., and C.E. Walsh (1991), "Testing intertemporal budget constraints: theory and applications to U.S. federal budget and current account deficits", *Journal of Money Credit and Banking*, 23, 206-23.

Wilcox, D.W. (1989), "The sustainability of government deficits: implications of the present value borrowing constraint", *Journal of Money, Credit and Banking*, 21, 291-306.

Appendix

A1. Generation of joint density forecast of linear and non-linear VECM using stochastic simulation

The stochastic simulation method explained in Galvao (2006) is used to produce the joint density forecasts. Define x_t as the vector of endogenous variables $\{\Delta G, \Delta R\}'$, $X^t = \{x_{t-1}, x_{t-2}, .., x_1\}$ as the history at time t . Given an estimate of $\hat{A}$ from the linear VECM model $x_t = f(X^{t-1}; A) + u_t$, and of the sample covariance matrix of residuals, $\hat{\Sigma}$, a trial sequence of forecasts $x_{t+1}, x_{t+2}, x_{t+3}, ..., x_{t+h}$ is built as follows. A random vector u_{t+1} is drawn from the distribution $u \sim N\left(0, \hat{\Sigma}\right)$ and it is used to calculate $\hat{x}_{t+1}$, given X^t and $\hat{A}$. Then $\hat{x}_{t+1}$ is added to "history" to form $\hat{X}^{t+1}$. This procedure is continued until the sequence of forecast is complete $\{x_{t+1}, x_{t+2}, x_{t+3}, ..., x_{t+h}\}$. This sequence of forecast can be called S_1, and the same trial is repeated to obtain a set of 1000 forecast sequences. In the case of threshold models, the forecasting model can be also written as $x_t^j = f^j(X^{t-1}; \beta^j) + \varepsilon_t^j$, where $j = 1,2$, to indicate the two regimes. Therefore, given $\hat{\Sigma}^1$ and $\hat{\Sigma}^2$, which are the estimated covariances for the two regimes, in order to obtain the forecast sequence we proceed as follows. Given the one step ahead point forecast, either the vector u_{t+h}^1 is drawn from $u^1 \sim N\left(0, \hat{\Sigma}^1\right)$ or the vector u_{t+h}^2 is drawn from $u^2 \sim N\left(0, \hat{\Sigma}^2\right)$, depending on whether the deficit is below or above the estimated threshold.

The realizations for this vector of innovations are then used to calculate $\hat{x}_{t+1}$, given X^t and $\hat{A}$. Then $\hat{x}_{t+1}$ is added to "history" to form $\hat{X}^{t+1}$. This procedure is continued until the sequence of forecast is complete $\{x_{t+1}, x_{t+2}, x_{t+3}, ..., x_{t+h}\}$. This sequence of forecast can be called S_1, and the same trial is repeated to obtain a set of 1000 forecast sequences. For each sequence of forecasts S_m, (with m describing the m^{th} scenario) we pick the last vector of observations, e.g. $x_{t+h,}$. The first component of this vector describes the joint model prediction for the (change in the) government spending series associated with scenario m and the second component of this vector describes the joint model prediction for the (change in the) tax revenue series associated with scenario m.

A2. Generation of conditional density forecast linear and non-linear VECM using stochastic simulation

The methodology to generate the sequence of forecast *S* (by picking the last observation in this sequence) is similar to the method described in A.1. The only exception consists of fixing to a specific value one of the two innovations, and this gives the conditional density forecast. In particular, if we fix the innovation to tax revenues to the sample mean of this series, and if we let the other shock (e.g. the one affecting government spending) to get 1000 realisations from Gaussian random draws, then we are able to generate the density forecast of government spending conditional on the sample mean value of tax revenues. Furthermore, if we fix the innovation to government spending to the sample mean of this series, and if we let the other shock (e.g. the one affecting tax revenues) to get 1000 realisations from Gaussian random draws, then we are able to generate the density forecast of tax revenues spending conditional on the sample mean value of tax revenues..

A.3 Generation of marginal density forecast of linear and non-linear VECM using stochastic simulation

The methodology to generate the sequence of forecast *S* (by picking the last observation in this sequence) is similar to the method described in A.1. However, the simulation method involves calibration to the sample standard deviation of each series and not to the overall sample covariance matrix. Specifically, the only difference with the method described in A.1 consists of multiplying the different realization of an iid shock (using standardized Gaussian random draws) by the sample standard deviation of government spending, thereby obtaining the marginal density forecast of government If we multiply the different realization of an iid shock (using standardized Gaussian random draws) by the sample standard deviation of tax revenues, thereby obtaining the marginal density forecast of tax revenues.

A.4 Generation of density forecast of a univariate AR model using stochastic simulation

Given the estimation of an AR(1) for each of the two series, the density forecasts at different horizon for one series is given by $x_{t+h} = \hat{\alpha}_0 h + (\hat{\alpha}_1^{h} x_t + \hat{\alpha}_1^{h-1} u_{t+1} + ... + u_{t+h})$, where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the estimated intercept and autoregressive coefficient of each series.

Table 1 Unit root tests on the level of the series R and G

	ADF	*PP*	ADF^{GLS}	MZ^{GLS}_{a}
R	-0.339	0.534	0.960	1.051
G	0.498	-0.514	2.053	1.633

Notes: In the first two columns we report the ADF and Philips-Perron test statistics, based upon an optimal lag selection through the BIC criterion (similar results are obtained using the AIC criterion). The critical values for both the ADF and Philips-Perron tests are from MacKinnon (1996) and they are equal to –3.46, -2.87, -2.57 for the 1%, 5%, 10% level of significance, respectively. In the last two columns we report the ADF^{GLS} and MZ^{GLS}_{a} statistics (for the case of only a constant in the deterministic component) developed by Ng and Perron (2001). The optimal lag selection has been carried using the modified AIC criterion suggested by Ng and Perron (2001). The 1%, 5% and 10% critical values for the MZ^{GLS}_{a} test are -13.8,-8.1 and -5.7%, respectively. The 1%,5% and 10% critical values for the ADF^{GLS} tests are -2.58,-1.98 and -1.62%, respectively.

Table 2a: Lag order for Linear VECM

Lag order	**AIC**	**BIC**
1	-9.781	-6.887
2	-9.474	-5.156
3	-6.78	-1.055
4	-0.409	6.711

Table 2b: Lag order for Threshold VECM

Lag order	**AIC**	**BIC**
1	-12.87	-7.082
2	-11.24	-2.609
3	-6.782	-1.055
4	7.798	22.04

Table 3- Tests for Threshold Cointegration

β = 1	
Lagrange Multiplier Threshold Test Statistic	18.700
Fixed Regressor Asymptotic p-Value	0.062
Bootstrap p-Value	0.085
Wald Test for Equality of	
Dynamic coeff	ECM coeff
Wald test = 23.19 *p-value* = 0.000	Wald-test = 6.120 *p-value* = 0.046

Notes: The p-values for the LM Threshold test were obtained by 5000 bootstrap replications. As for the Wald test, p-values are in the parenthesis.

Table 4 - Estimates of the Threshold VAR

	$\beta = 1$			
	Threshold Estimate = 8.859			
	Regime 1		**Regime 2**	
	ΔG	ΔR	ΔG	ΔR
Intercept	0.312 (0.066)	0.068 (0.101)	3.137 (1.160)	0.988 (1.336)
w_{t-1}	-0.010 (0.014)	0.023 (0.022)	-0.242 (0.096)	-0.012 (0.113)
ΔG_{t-1}	-0.216 (0.135)	-0.040 (0.099)	-0.142 (0.125)	-0.514 (0.246)
ΔR_{t-1}	-0.094 (0.055)	0.101 (0.138)	-0.088 (0.084)	-0.697 (0.153)
% of observations in regime	82%		18%	

Notes: Standard errors in parentheses.

Table 5: RMSE for point forecast

forecast horizon	AR		Linear VECM		Threshold VECM	
	ΔG	***ΔR***	***ΔG***	***ΔR***	***ΔG***	***ΔR***
H = 1	0.823	1.507	0.824	1.517	0.821	1.781
H = 4	0.808	1.503	0.774	1.465	0.784	1.484
h = 8	0.795	1.509	0.807	1.462	0.827	1.514

Notes: The RMSE associated with the point forecasts have been obtained by recursive estimation of both Linear and Non-Linear VECM, using the sample running from 1989:1 to 2004:4 as the forecast evaluation period.

Table 6a: LM test for iid of probability integral transform

1 quarter ahead forecasts						
Moment	AR		Linear VECM		Threshold VECM	
	ΔG	***ΔR***	***ΔG***	***ΔR***	***ΔG***	***ΔR***
1	0.143	0.002	0.127	0.006	0.130	0.014
2	0.560	0.284	0.510	0.104	0.463	0.132
3	0.228	0.051	0.209	0.051	0.189	0.077
			ΔG/ΔR	***ΔR/ΔG***	***ΔG/ΔR***	***ΔR/ΔG***
1			0.126	0.008	0.126	0.015
2			0.515	0.120	0.469	0.138
3			0.211	0.076	0.199	0.0824
			(ΔG/ΔR) *ΔR	***(ΔR/ΔG) *ΔG***	***(ΔG/ΔR) *ΔR***	***(ΔR/ΔG) *ΔG***
1			0.122	0.006	0.126	0.003
2			0.500	0.112	0.548	0.228
3			0.211	0.0753	0.185	0.057

Note: The table records the p-values for χ^2 LM tests of serial correlation (up to fourth order) for the first, second and third moments of the probability integral transform, PIT, series.

Table 6b: LM test for iid of probability integral transform

4 quarters ahead forecasts						
Moment	AR		Linear VECM		Threshold VECM	
	ΔG	***ΔR***	***ΔG***	***ΔR***	***ΔG***	***ΔR***
1	0.147	0.001	0.174	0.004	0.097	0.012
2	0.465	0.326	0.544	0.199	0.592	0.119
3	0.362	0.093	0.264	0.050	0.148	0.075
			ΔG/ΔR	***ΔR/ΔG***	***ΔG/ΔR***	***ΔR/ΔG***
1			0.162	0.0103	0.108	0.007
2			0.624	0.161	0.657	0.095
3			0.245	0.116	0.188	0.0304
			(ΔG/ΔR) *ΔR	***(ΔR/ΔG) *ΔG***	***(ΔG/ΔR) *ΔR***	***(ΔR/ΔG) *ΔG***
1			0.154	0.006	0.095	0.011
2			0.539	0.108	0.625	0.105
3			0.249	0.0593	0.147	0.053

Note: Note: see Note to Table 6a

Table6c: LM test for iid of probability integral transform

8 quarters ahead forecasts						
Moment	**AR**		**Linear VECM**		**Threshold VECM**	
	ΔG	***ΔR***	***ΔG***	***ΔR***	***ΔG***	***ΔR***
1	0.134	0.002	0.165	0.006	0.121	0.008
2	0.459	0.301	0.568	0.152	0.526	0.081
3	0.319	0.099	0.272	0.058	0.181	0.057
			ΔG/ΔR	***ΔR/ΔG***	***ΔG/ΔR***	***ΔR/ΔG***
1			0.172	0.008	0.107	0.009
2			0.606	0.175	0.592	0.063
3			0.265	0.089	0.175	0.028
			(ΔG/ΔR) *ΔR	***(ΔR/ΔG) *ΔG***	***(ΔG/ΔR) *ΔR***	***(ΔR/ΔG) *ΔG***
1			0.168	0.008	0.0942	0.007
2			0.505	0.147	0.512	0.097
3			0.251	0.0656	0.137	0.041

Note: Note: see Note to Table 6a

Table 7a: Berkowitz test for 1 quarter ahead forecasts

AR		Linear VECM		Threshold VECM	
ΔG	ΔR	ΔG	ΔR	ΔG	ΔR
0.000	0.000	0.000	0.000	0.000	0.000
		$\Delta G/\Delta R$	$\Delta R/\Delta G$	$\Delta G/\Delta R$	$\Delta R/\Delta G$
		0.000	0.000	0.072	0.000
		$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$	$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$
		0.000	0.000	0.104	0.000

Note: The entries are the p-values of the Berkowitz (1999) Likelihood ratio test for joint null of normality and iid in PIT*, which is the inverse of the cumulative normal distribution of the probability integral transform, PIT.

Table 7b: Berkowitz test for 4 quarter ahead forecasts

AR		Linear VECM		Threshold VECM	
ΔG	ΔR	ΔG	ΔR	ΔG	ΔR
0.013	0.000	0.000	0.000	0.010	0.000
		$\Delta G/\Delta R$	$\Delta R/\Delta G$	$\Delta G/\Delta R$	$\Delta R/\Delta G$
		0.000	0.000	0.013	0.000
		$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$	$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$
		0.000	0.000	0.012	0.000

Note: see Note to Table 7a

Table 7c: Berkowitz test for 8 quarter ahead forecasts

AR		Linear VECM		Threshold VECM	
ΔG	ΔR	ΔG	ΔR	ΔG	ΔR
0.000	0.000	0.000	0.000	0.074	0.000
		$\Delta G/\Delta R$	$\Delta R/\Delta G$	$\Delta G/\Delta R$	$\Delta R/\Delta G$
		0.000	0.000	0.031	0.000
		$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$	$(\Delta G/\Delta R)*\Delta R$	$(\Delta R/\Delta G)*\Delta G$
		0.000	0.000	0.048	0.000

Note: see Note to Table 7a

Table 8a: Probability forecast evaluation for government spending

	QPS	LPS
AR	0.467 0.514 0.497	0.660 0.707 0.690
Linear VECM	0.473 0.443 0.441	0.664 0.635 0.633
Non-Linear VECM	0.464 0.434 0.440	0.652 0.626 0.632

Note: The three entries in each cell (from the top to the bottom) are the QPS and LPS scores for the one, four and eight quarters ahead

Table 8b: Probability forecast evaluation for tax revenues

	QPS	LPS
AR	0.478 0.505 0.499	0.671 0.698 0.692
Linear VECM	0.496 0.478 0.479	0.692 0.671 0.672
Non-Linear VECM	0.584 0.481 0.470	0.873 0.678 0.662

Note: see Note to Table 8

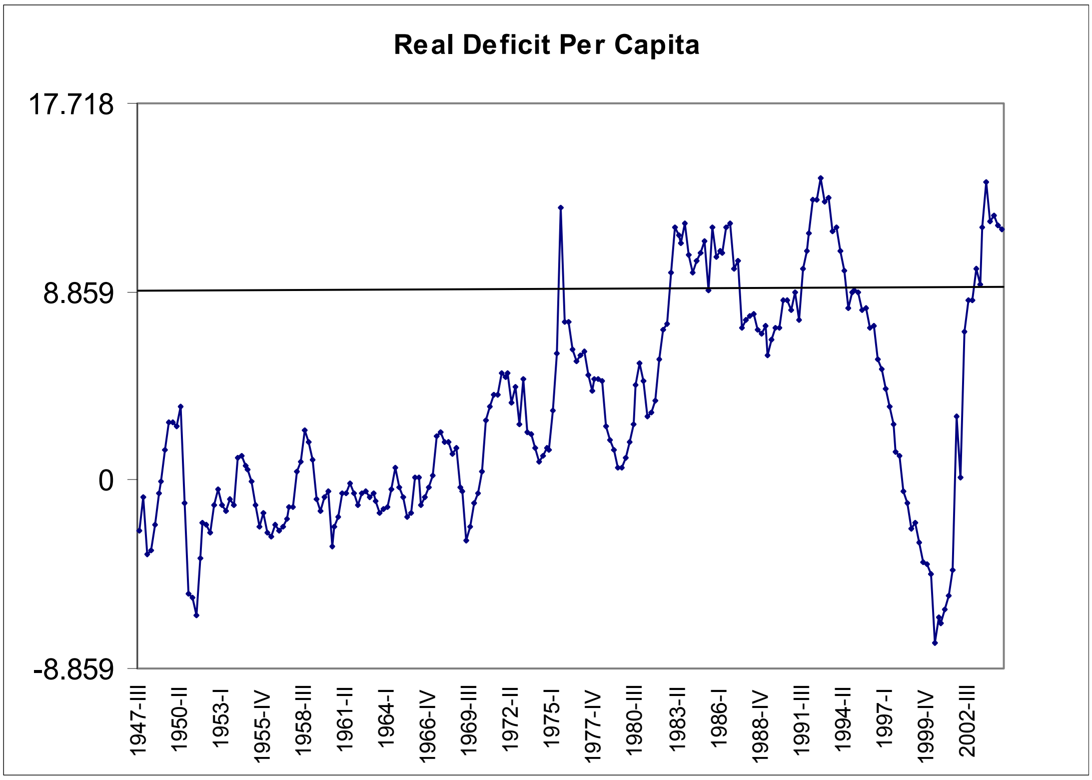
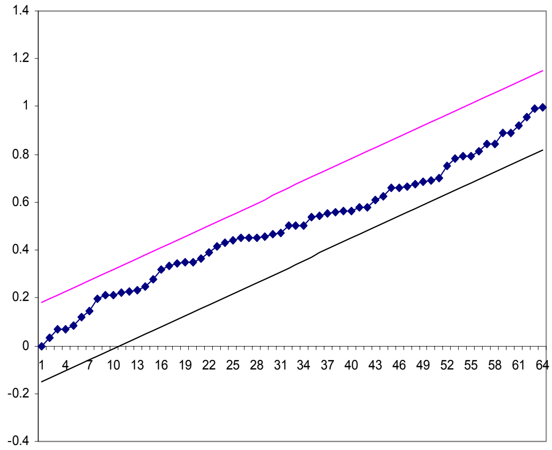


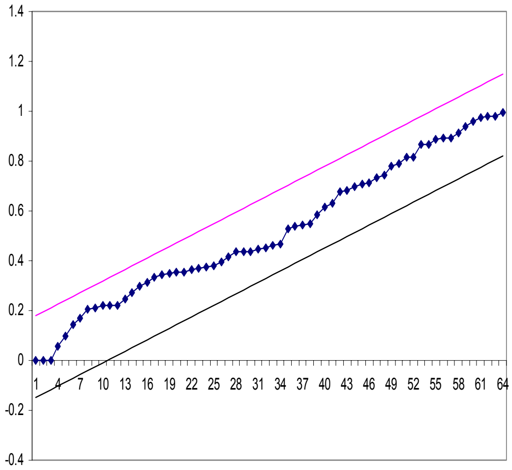
Figure 1: Identification of Threshold Regimes

Figure 2: PP plots for PIT corresponding to AR density forecasts

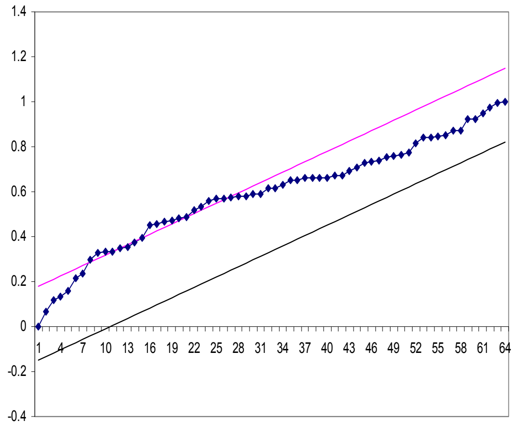
ΔG: 1 step ahead



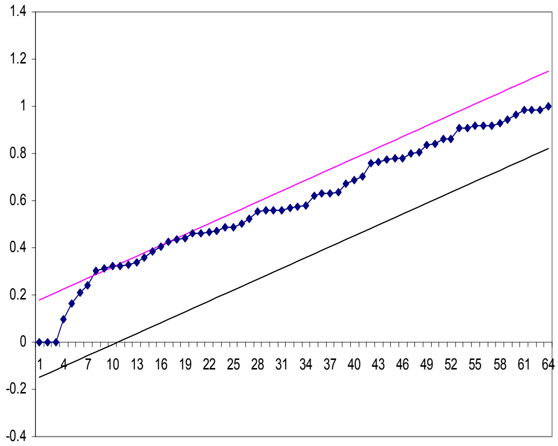
ΔR: 1 step ahead



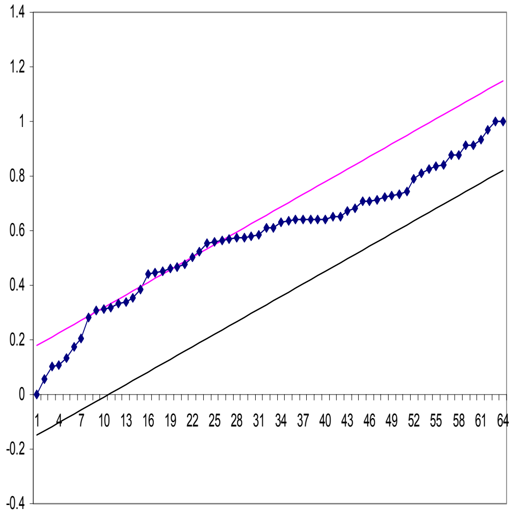
ΔG: 4 step ahead



ΔR: 4 step ahead



ΔG: 8 step ahead



ΔR: 8 step ahead

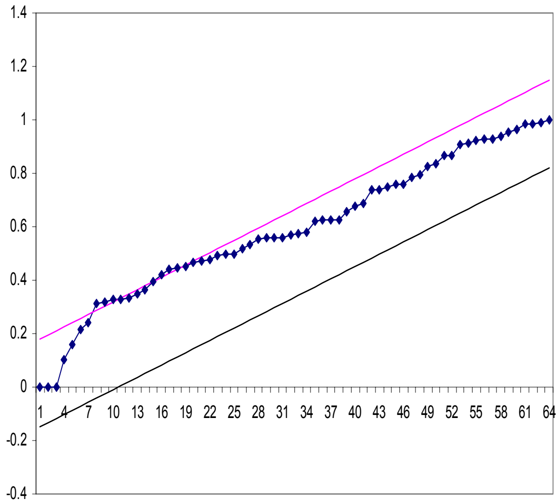
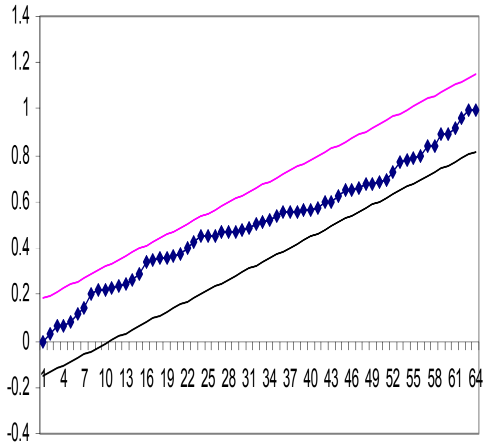
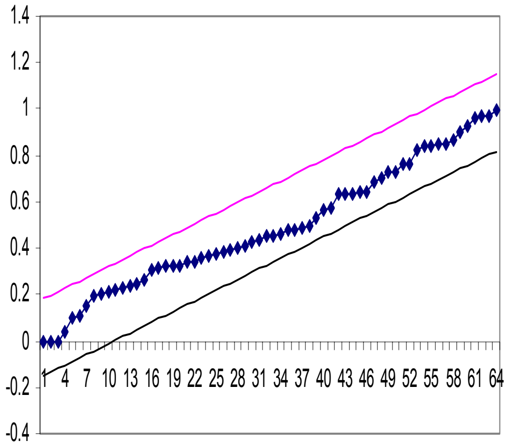


Figure 3: PP plots for PIT corresponding to linear VECM marginal density forecasts

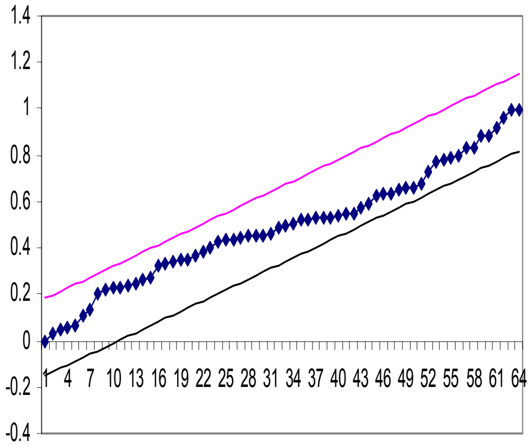
ΔG: 1 step ahead



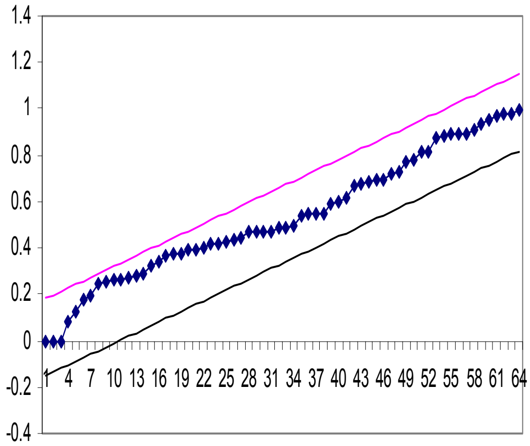
ΔR: 1 step ahead



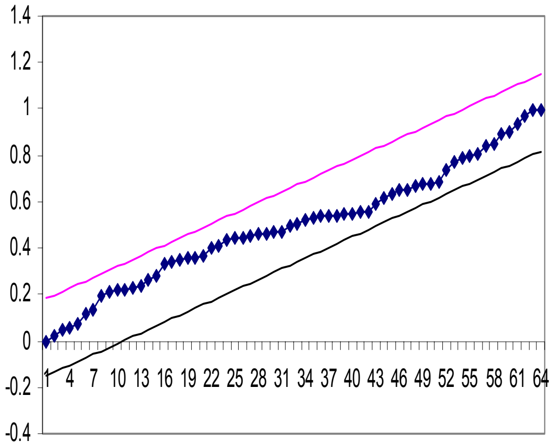
ΔG: 4 step ahead



ΔR: 4 step ahead



ΔG: 8 step ahead



ΔR: 8 step ahead

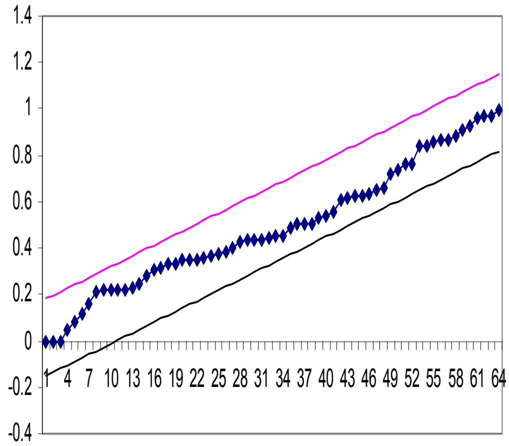
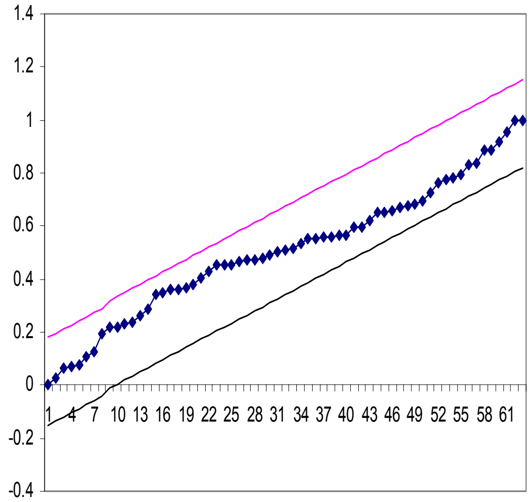
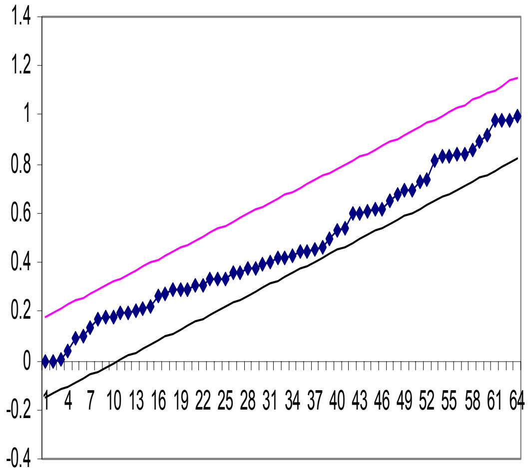


Figure 4: PP plots for PIT corresponding to linear VECM conditional density forecasts

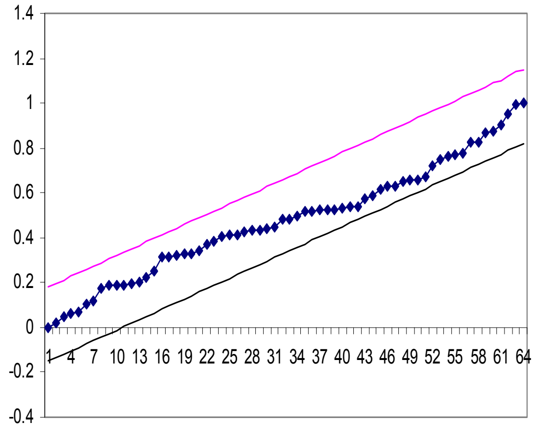
ΔG/ΔR: 1 step ahead



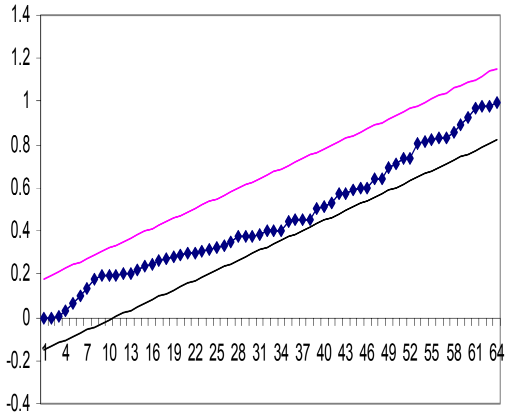
ΔR/ΔG: 1 step ahead



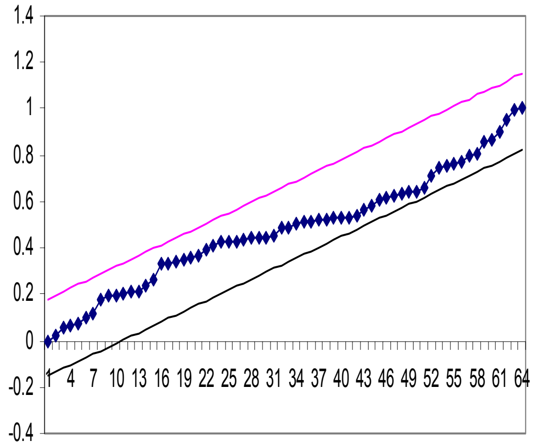
ΔG/ΔR: 4 step ahead



ΔR/ΔG: 4 step ahead



ΔG/ΔR: 8 step ahead



ΔR/ΔG: 8 step ahead

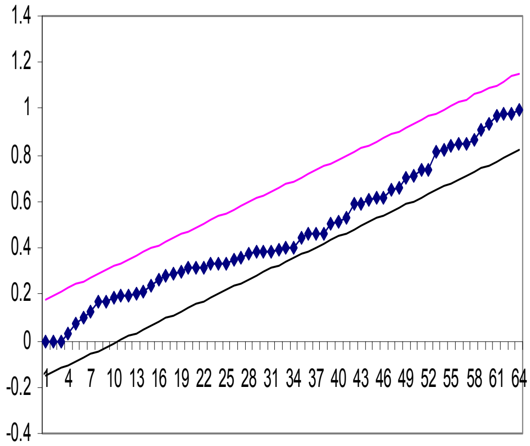
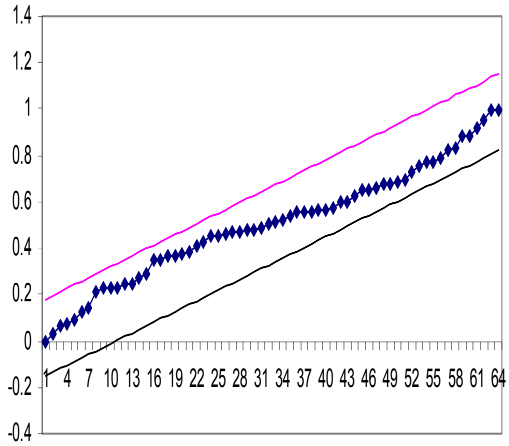
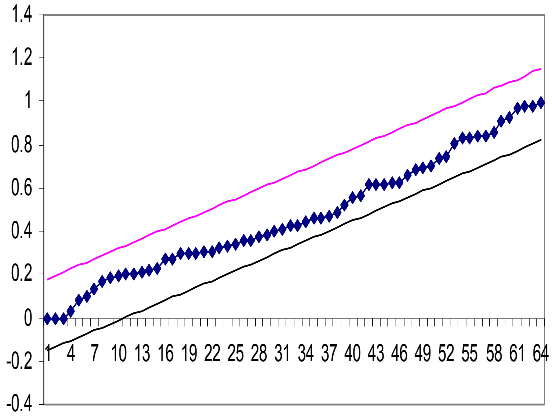


Figure 5: PP plots for PIT corresponding to linear VECM joint density forecasts

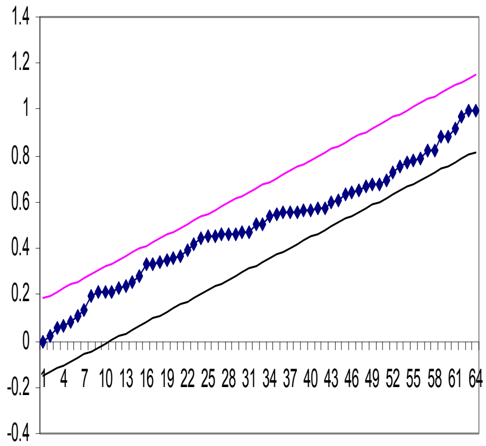
(ΔG/ΔR)* ΔR: 1 step ahead



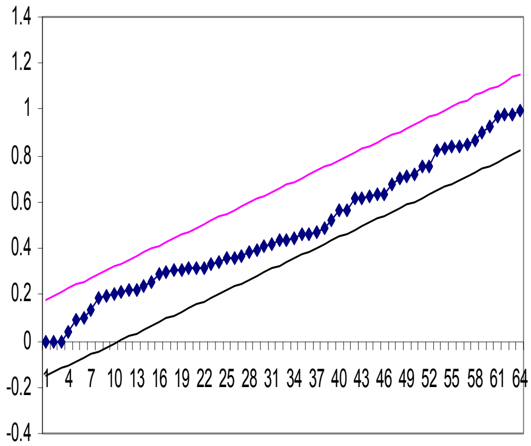
(ΔR/ΔG)* ΔG: 1 step ahead



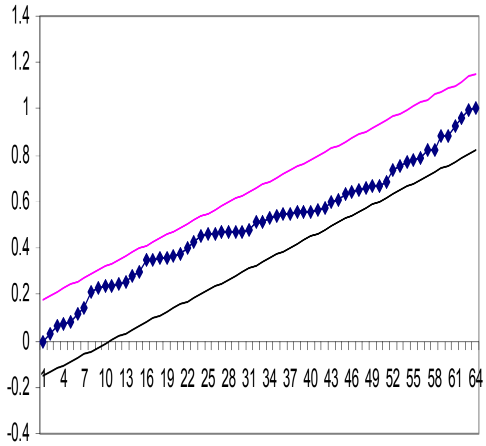
(ΔG/ΔR)* ΔR: 4 step ahead



(ΔR/ΔG)* ΔG: 4 step ahead



(ΔG/ΔR)* ΔR: 8 step ahead



(ΔR/ΔG)* ΔG: 1 step ahead

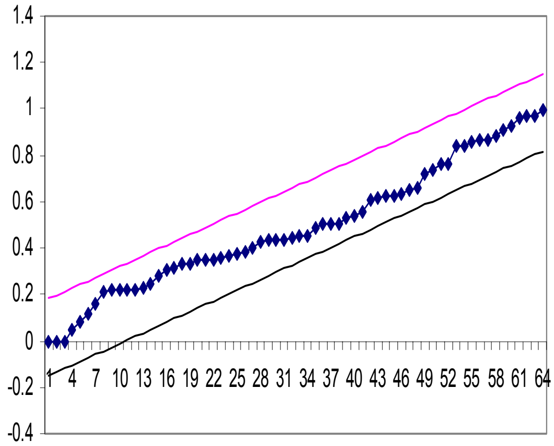
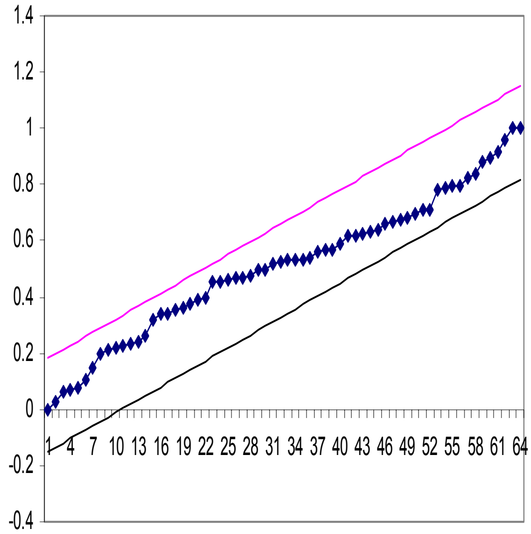
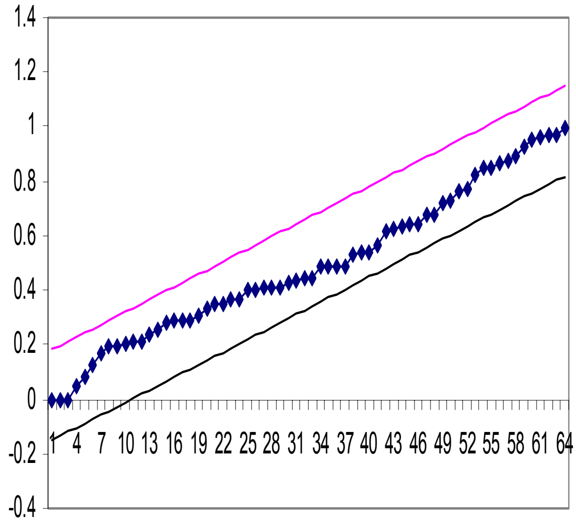


Figure 6: PP plots for PIT corresponding to threshold VECM marginal density forecasts

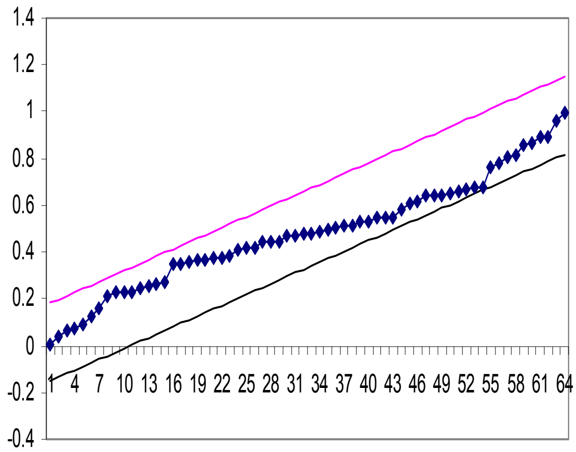
ΔG: 1 step ahead



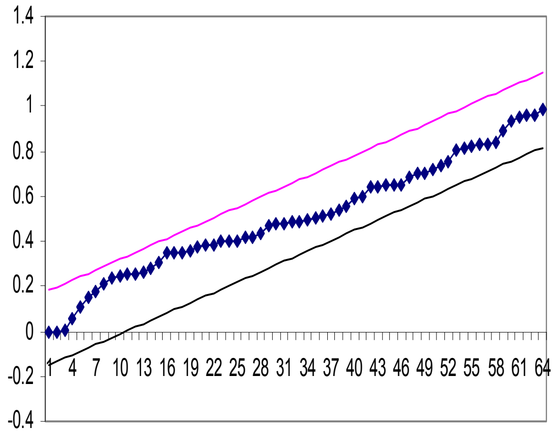
ΔR: 1 step ahead



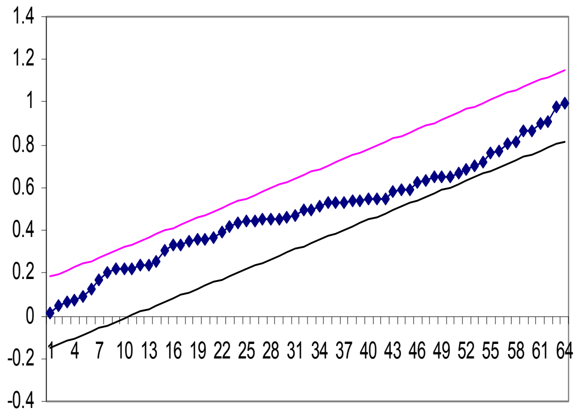
ΔG: 4 step ahead



ΔR: 4 step ahead



ΔG: 8 step ahead



ΔR: 8 step ahead

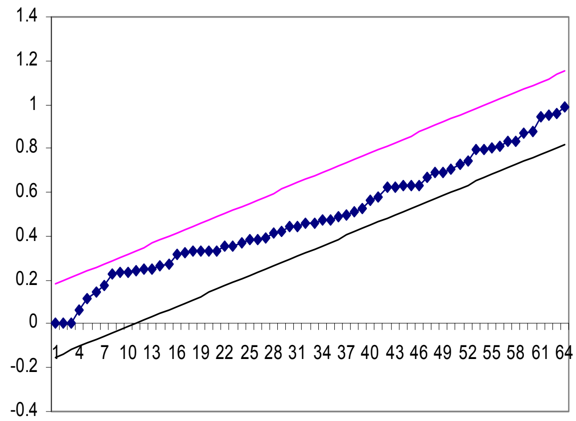
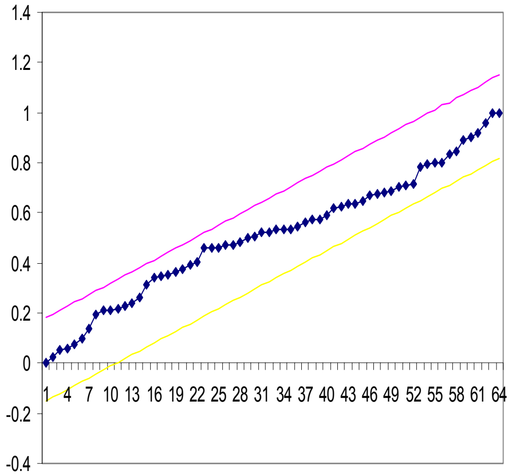
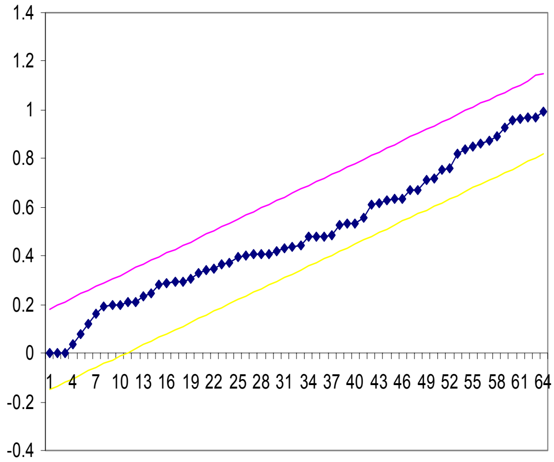


Figure 7: PP plots for PIT corresponding to threshold VECM conditional density forecasts

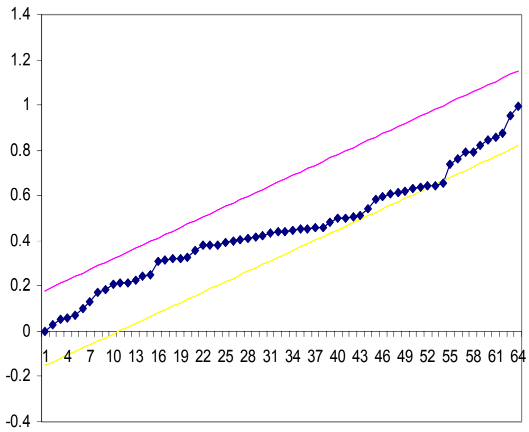
ΔG/ΔR: 1 step ahead



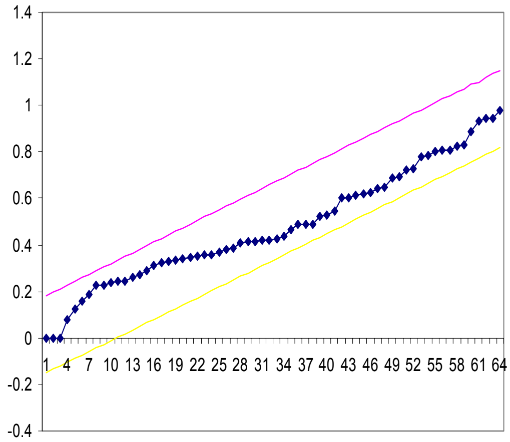
ΔR/ΔG: 1 step ahead



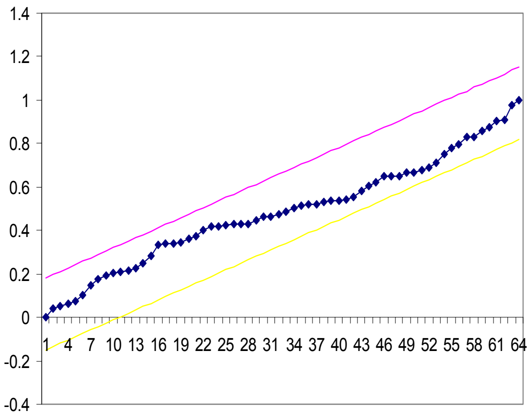
ΔG/ΔR: 4 step ahead



ΔR/ΔG: 4 step ahead



ΔG/ΔR: 8 step ahead



ΔR/ΔG: 8 step ahead

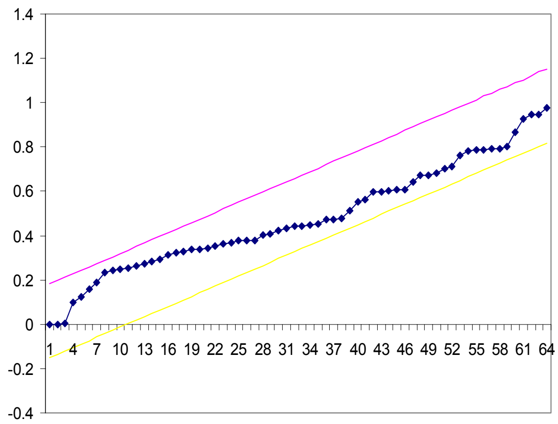
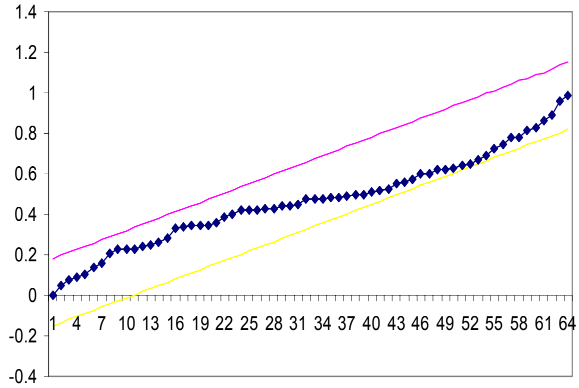
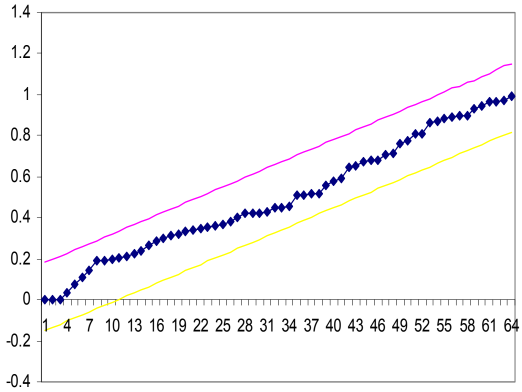


Figure 8: PP plots for PIT corresponding to threshold VECM joint density forecasts

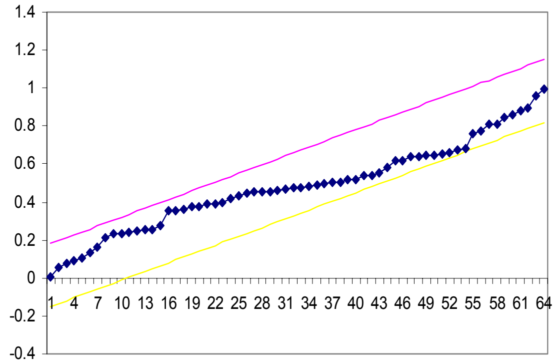
(ΔG/ΔR)* ΔR: 1 step ahead



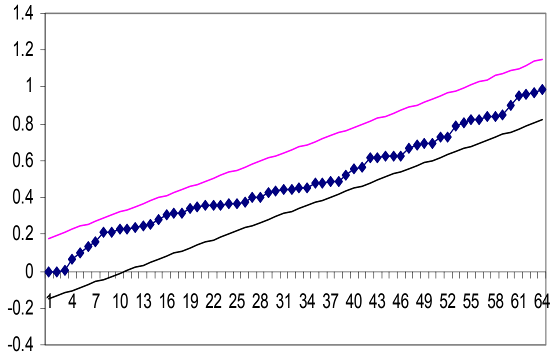
(ΔR/ΔG)* ΔG: 1 step ahead



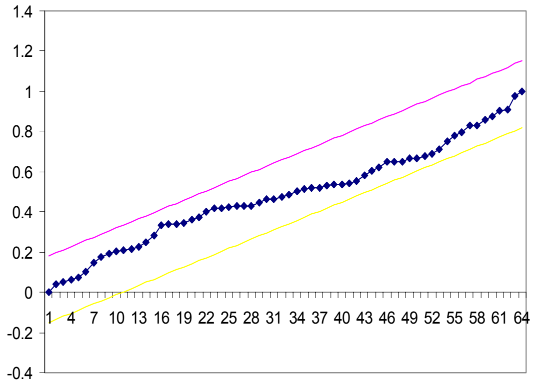
(ΔG/ΔR)* ΔR: 4 step ahead



(ΔR/ΔG)* ΔG: 4 step ahead



(ΔG/ΔR)* ΔR: 8 step ahead



(ΔR/ΔG)* ΔG: 8 step ahead