

# Instrumental variable estimation of a nonlinear Taylor rule

Zisimos Koustas<sup>a</sup>, Jean-François Lamarche<sup>a</sup>

<sup>a</sup> Department of Economics, Brock University, St. Catharines, Ontario, Canada

December 18, 2009

## Abstract

This paper estimates a nonlinear threshold model using instrumental variables. This estimation strategy was originally developed with dynamic panel models in mind and we extend it to time series models. In particular, we consider a forward-looking Taylor rule and test to see if the Bank of England followed a nonlinear Taylor rule in setting the short-term interest rate.

*Keywords: Thresholds, Nonlinear Models, Instrumental Variables, Taylor Rule*

*JEL codes: C22, C12, C13, C87, E58*

Corresponding Author:  
Jean-François Lamarche  
Department of Economics, Brock University  
500 Glenridge Avenue, St. Catharines  
Ontario, Canada, L2S 3A1  
Phone: 905-688-5550 ext. 3328  
E-mail: jfl@brocku.ca

## 1 Introduction

John Taylor (see Taylor (1993)) - formalized the notion that the US Federal Reserve sets the target for the federal funds rate in order to establish the appropriate level of the real interest rate, with adjustments aimed at correcting deviations of the rate of inflation from its target and of output from potential. A considerable body of theoretical work on optimal monetary policy rules has since emerged and has been accompanied by numerous empirical studies testing for the use of such rules by central banks.

Much of the empirical literature on the Taylor rule has focused on the US Federal Reserve - see for example Clarida, Galí, and Gertler (2000) and Orphanides (2001) and (2003). In an international setting, Taylor (1999) studied different strategies for setting interest rates by the European Central Bank; Clarida, Galí, and Gertler (1998) studied the Taylor rule for the UK, France, Germany, Italy, Japan and the US; Nelson (2001) focused solely on the UK, and more recently in an interesting paper Taylor and Davradakis (2006) follow the framework of Clarida, Galí, and Gertler (1998) and estimate a forward-looking version of the Taylor rule for the UK using the generalized method of moments (GMM). Their approach is novel and challenging in that they consider a nonlinear version of the Taylor rule that incorporates the idea of a threshold autoregression in which the interest rate follows different regimes characterized by an inflation threshold. Hence, depending on the severity of inflation, the central bank may resort to different types of monetary policy. For example, if inflation is above a certain level, a policy combatting inflation aggressively might be implemented. Therefore, even if a central bank has a statement relating to a specific inflation target, it is possible to see some degree of nonlinearity in the bank's policy setting. Bunzel and Enders (2005) have also looked at this possibility using a backward looking version of the Taylor rule for the US.

The purpose of this paper is to see whether monetary policy could be characterized by a rule that is linear in its key policy tool (inflation, for example) or rather by a rule that depends on different thresholds that inflation can take. That is, we want to test the null hypothesis of a linear Taylor rule versus the alternative of a nonlinear rule. The estimation and inference framework for linear models with exogenous regressors is well established (see for example Chan (1993), Hansen (1996) and Hansen (2000)). When some of the regressors are endogenous, care must be taken when estimating the parameters and in the way the threshold estimate is obtained. For this, we follow the approach of Caner and Hansen (2004), hereafter CH.

We apply the multi-step estimation strategy proposed by CH using the data of Taylor and Davradakis (2006), hereafter TD. A by-product of this exercise relates to an appropriate estimation method. Instead of using an instrumental variables approach under the assumptions of homogeneity and no serial correla-

tion (yielding sub-optimal estimators), CH allowed for heteroskedasticity. This is typically accomplished through the use of consistent but inefficient estimators (generalized instrumental variable estimators, for example) in a first-stage regression. The covariance matrix estimated in the first-stage is then used in a second-stage to obtain efficient estimators in the presence of serial correlation and/or heteroskedasticity. CH have shown the correct use of efficient GMM in the context of a threshold model. In a footnote we also illustrate that practitioners must be careful in their choice and/or in their use of software.

The main findings are summarized as follows. We find that the null hypothesis of a linear model is not rejected, implying that, if the Bank of England has followed a Taylor rule, it is linear with precise parameter estimates. The underlying key parameters indicate that the Bank was concerned with the output gap, as well as, with inflation during the 1992-2003 period. This seems to be in accord with the explicit inflation-targeting regime that was initiated in 1992.

This paper is organized as follows. The second section presents two versions of the Taylor rule, the third section discusses the estimation methods and threshold selection, the fourth section summarizes the results and the last section concludes.

## 2 A linear and a nonlinear Taylor rule

Following Clarida, Galí and Gertler (2000) we use a policy rule for the target Federal Funds rate  $i_t^*$  as

$$i_t^* = i^* + \delta(E(\pi_{t,k}|\Omega_t) - \pi^*) + \gamma E(x_{t,q}|\Omega_t) \quad (1)$$

where  $i^*$  is the target interest rate when expected inflation is equal to its target ( $\pi^*$ ) and the output gap is zero.  $\pi_{t,k}$  is the inflation rate between periods  $t$  and  $t+k$ ,  $x_{t,q}$  is the output gap  $q$  periods into the future and  $\Omega_t$  is the information set on which expectations are based. Subtracting expected inflation from both sides, and defining  $ii^* = i^* - \pi^*$  as the real interest rate target when expected inflation is equal to the target and the output gap is zero, we get a rule of the real rate target

$$i_t^* - E(\pi_{t,k}|\Omega_t) = ii^* + (\delta - 1)(E(\pi_{t,k}|\Omega_t) - \pi^*) + \gamma E(x_{t,q}|\Omega_t). \quad (2)$$

When  $\delta > 1$  and  $\gamma > 0$  the policy rule is stabilizing while it is destabilizing if  $\delta < 1$  or  $\gamma < 0$ . Denoting the actual Funds rate by  $i_t$  and assuming that the Federal Reserve smoothes changes in its interest rate towards the target  $i_t^*$  we have a smoothing function

$$i_t = (1 - \rho)i_t^* + \rho i_{t-1} + v_t \quad (3)$$

where  $\rho$  is a parameter measuring the degree of smoothing of interest rate changes and  $v_t$  is an exogenous interest shock with 0 mean. Note that we could have a more complex smoothing function by adding more

lags of the actual rate (for example Clarida, Galí, and Gertler (2000) introduce enough lags so that the error term is uncorrelated). As  $\rho$  increases, changes in the rate will tend to be slow. Combining equations (1) and (3) we get the forward-looking, linear, Taylor rule<sup>1</sup> (this is model M4 in TD)

$$i_t = \alpha_0 + \alpha_1 i_{t-1} + \alpha_2 \pi_{t,k} + \alpha_3 x_{t,q} + u_t \quad (4)$$

where the underlying key parameters are given by  $\delta = \alpha_2/(1-\rho)$ ,  $\gamma = \alpha_3/(1-\rho)$ ,  $\rho = \alpha_1$  and an intercept of  $ii^* = \alpha_0/(1-\rho)$ . Note that with a given inflation target we can obtain an estimate of the long run real interest rate,  $ii^*$  (see Clarida, Galí, and Gertler (1998) for example).

Allowing for a nonlinear Taylor rule can be accomplished by distinguishing between different policy regimes depending on the level of inflation. Because the results of CH require the use of an exogenous threshold variable, lagged rather than contemporaneous inflation will be used as the threshold. This is a departure from the work of TD who used current inflation as the threshold. Andros, Stengos, and Tan (2009) do consider the case of an endogenous threshold, an avenue of research that is not explored in this paper. The simplest two-regime model (model M2 in TD) can then be written as:

$$i_t = (\alpha_0 + \alpha_1 i_{t-1} + \alpha_2 \pi_{t,k} + \alpha_3 x_{t,q}) 1(\pi_{t-1} \geq \lambda) + (\beta_0 + \beta_1 i_{t-1} + \beta_2 \pi_{t,k} + \beta_3 x_{t,q}) 1(\pi_{t-1} < \lambda) + u_t \quad (5)$$

where  $1(\cdot)$  is an indicator function and  $\lambda$  is the threshold value of inflation. Note that when  $\alpha = \beta$  (where  $\alpha$  and  $\beta$  are  $4 \times 1$  vectors) we have the linear Taylor rule (4). Estimation of the threshold is typically performed using a one-dimensional grid search over potential values of the threshold. In linear, non-IV, models the optimal value of the threshold is found by minimizing the residual variance of the model or equivalently the sum of squared residuals, from (5), over the set of possible values for  $\lambda$ . That is, letting  $\hat{\sigma}^2(\lambda) = (1/T) \sum_{t=1}^T \hat{u}_t(\lambda)^2$ , the following problem is solved

$$\hat{\lambda} = \min_{\lambda \in \Lambda} \hat{\sigma}^2(\lambda)$$

where  $\Lambda$  defines the set of possible values for the threshold.

### 3 Estimation method

#### 3.1 The linear case

Because the inflation rate,  $\pi_{t,k}$ , and the output gap,  $x_{t,q}$ , are multi-step ahead forecasts and we are using realized values for these forecasts, ordinary least squares (OLS) estimators will be inconsistent since

---

<sup>1</sup>The expectations operator is dropped for notational convenience.

the regressors (matrix  $X$ ) will be correlated with the error term, that is,

$$EX'u \neq 0 \tag{6}$$

and the condition

$$E(u_t|\Omega_t) = 0 \tag{7}$$

will be violated. Typically the problem is resolved with the use of instruments that lie in the information set  $\Omega_t$  but that are uncorrelated with the error term  $u_t$ . Essentially these instruments isolate, and ignore, the component of the regressors that is correlated with  $u_t$ . Without loss of generality we can rewrite the previous model as

$$y = X\theta + u \tag{8}$$

where  $y$  is a  $T \times 1$  vector containing  $i_t$ ,  $X$  is a  $T \times 4$  matrix containing a constant,  $i_{t-1}$ ,  $\pi_{t,k}$  and  $x_t$ ;  $\theta$  is a  $4 \times 1$  vector of unknown coefficients and  $u$  is a  $T \times 1$  vector of error terms. Although this section focuses on the linear Taylor rule, the nonlinear version of this rule is such that nonlinearity is introduced via thresholds. These thresholds give us regimes within which OLS can be applied. The discussion in this section, therefore, can be extended to the case of nonlinear models.

Denoting the instruments set by  $W$ , a  $T \times l$  matrix with  $l = 4$ , we have the following modified orthogonality (moment) conditions

$$EW'u = 0.$$

Clarida, Galí and Gertler (2000) and TD specified  $W$  as containing a constant and the first six lags plus the 9th and 12th lags of the interest rate, inflation, the producer price index and the output gap. They used as instruments three measures of inflation (one based on the retail price index, one based on this index less mortgage interest payments and a last one which also excludes indirect taxes). In this scenario a total of 49 instruments were used. Further an inflation target of 3 months was used (that is  $k = 3$ ). This is in contrast with Clarida, Galí and Gertler (1998) who used  $k = 12$  with monthly data and Clarida, Galí and Gertler (2000) with  $k = 3$  for quarterly data. In the empirical section we estimate the model for  $1 \leq k \leq 12$  and select the best fitting model. In most applications, emphasis is put on forecasts of inflation rather than on those for the output gap ( $q$  is set to 1). We experimented with various choices of  $q$  and found that this time horizon did not matter.

If there are at least as many instruments as there are parameters ( $l \geq 4$ ), then the generalized IV moment conditions become

$$X'P_W(y - X\theta) = 0$$

where  $P_W$  is a projection matrix,  $W(W'W)^{-1}W'$ , yielding

$$\hat{\theta}_{IV} = (X'P_WX)^{-1}X'P_Wy. \quad (9)$$

The generalized IV moment conditions can also be seen more explicitly as arising from the model

$$y = P_wX\theta + v \quad (10)$$

where we can think of  $P_wX$  as fitted values from a set of regressions of all variables on the instruments (this would be the first stage of two-stage least squares).

To see the link between IV estimation and GMM, both in terms of estimation and specification testing, we note that this estimator can also be obtained by minimizing the IV criterion function

$$Q(\theta, y) = (y - X\theta)'P_W(y - X\theta) \quad (11)$$

or

$$Q(\theta, y) = (y - X\theta)'W(W'W)^{-1}W'(y - X\theta)$$

and has variance

$$Var(\hat{\theta}_{IV}) = \hat{\sigma}^2(X'P_WX)^{-1}$$

where  $\hat{\sigma}^2 = (y - X\hat{\theta}_{IV})'(y - X\hat{\theta}_{IV})/T$ . Now since we have more instruments than regressors we may want to see if the additional restrictions are necessary. The test proposed by Sargan (1958) is:

$$\frac{Q(\hat{\theta}_{IV}, y)}{\hat{\sigma}^2}$$

and it is asymptotically distributed as  $\chi^2(l - k)$ .

The efficient GMM estimator can be obtained via the minimization of the GMM criterion function (see Davidson and MacKinnon (2003))

$$Q(\theta, y) = (y - X\theta)'W(W'\Omega W)^{-1}W'(y - X\theta) \quad (12)$$

where from now on,  $\Omega$  is defined implicitly by  $E(uu') = \Omega$  and it is therefore the covariance matrix of the error terms. Note that if homogeneity and no serial correlation are assumed, we have  $\Omega = \sigma^2I$ ;  $I$  is an identity matrix, and the GMM criterion function is just the IV function (11) divided by  $\sigma^2$ . This is the approach taken by TD, yielding sub-optimal estimators in the event that the assumptions of homogeneity and no serial correlation are invalid. Note also that the GMM criterion function can be used directly to test for overidentifying restrictions. If  $\Omega$  is known, the minimization problem is straightforward. If, on the

other hand,  $\Omega$  is unknown it must be estimated. For example, if we have heteroskedasticity of unknown form, we can use the results of White (1980) and replace  $(1/T)W'\Omega W$  by  $(1/T)W'\hat{\Omega}W$  where  $\hat{\Omega}$  is a diagonal matrix containing elements of the form  $\hat{u}_t^2$ . Caner and Hansen (2004) use White's approach in their estimation method. When this is done we obtain the feasible efficient, two-step, GMM estimator

$$\theta = (X'W(W'\hat{\Omega}W)^{-1}W'X)^{-1}X'W(W'\hat{\Omega}W)^{-1}W'y. \quad (13)$$

The vector of residuals  $\hat{u}$  must first be computed. This can be accomplished using the generalized IV estimator specified above. As Hansen, Heaton, and Yaron (1996) have shown, this process can be iterated to obtain the continuously updated GMM estimator. Since both Clarida, Galí and Gertler (2000) and TD use a two-step procedure our analysis will focus on the latter.

In time series application the matrix  $\Omega$  may very well contain non-zero off-diagonal elements to capture serial correlation. In the present context, because multi-step forecasts are present in the Taylor rule, (4), the error term will be serially correlated (see Clarida, Galí and Gertler (2000) and Inoue and Shintani (2006)). In this case, we replace  $(1/T)W'\Omega W$  by the heteroskedasticity and autocorrelation consistent (HAC) estimator

$$\hat{\Sigma} = \frac{1}{T}W'\hat{\Omega}W \quad (14)$$

which requires an estimator of  $\Sigma$  given by

$$\Sigma = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E(u_t u_s W_t' W_s) \quad (15)$$

and more specifically an estimator of the autocovariance matrices of  $W_t' u_t$ . One popular choice is the Newey-West estimator (see Newey and West (1987) and (1994)). Estimation then proceeds just as before to obtain

$$\hat{\theta} = (X'W\hat{\Sigma}^{-1}W'X)^{-1}X'W\hat{\Sigma}^{-1}W'y \quad (16)$$

where in a first-stage we compute  $\hat{u}$  using the IV estimator (9).

Given the preceding discussion, care should be taken when using built-in functions in statistical software.<sup>2</sup>

---

<sup>2</sup>This footnote compares the parameter estimates of model M4, M4 is the linear Taylor rule (4) in TD, obtained using two different software packages, and illustrate that the results are not sensitive to the choice of software. It is shown, however, that the results are sensitive to the choice of estimators. Specifically, the use of the software package with built-in functions leads to sub-optimal estimators unless the appropriate command options are used. The first package is RATS (see Estima (2006)), an excellent and widely used software with many built-in functions, and the second is Ox (see Doornik (2007)), a matrix oriented programming language. The computer code is in the Appendix. We use the two-step estimation procedure as suggested by TD with a HAC estimator of  $\Sigma$ . In RATS this is implemented with the function *mcov* with the bandwidth set to 2. In Ox this is accomplished with the procedure *nw* (line 7). The RATS code makes use of some built-in functions. Part A of the code produces appropriate results for the two-step procedure. In part A, two-step estimation is performed with *nlsystem* a built-in function which targets multivariate, nonlinear optimization. Part B of the RATS code is the same

### 3.2 The nonlinear case

With a nonlinear Taylor rule, and with a given threshold that splits the sample into two parts, the estimation strategy is identical to the one described above with GMM, (16), applied to each subsample. We follow the work of CH who proposed a sequential estimation method for the unknown threshold. In a first stage, a regression of all endogenous variables on the matrix of instruments is performed to obtain the term  $P_w X$  in (10). The threshold is then estimated in a second regression (analogous to the second stage of a 2SLS procedure) that uses the splitted predicted values from the first stage as independent variables. Lastly GMM estimation is performed on each subsample. This can be summarized as:

1. Obtain the fitted values  $P_w X$ .
2. Letting the residuals,  $v_t$ , from (10) be defined implicitly by

$$y_t = (P_w X)_t \alpha 1(\pi_{t-1} \geq \lambda) + (P_w X)_t \beta 1(\pi_{t-1} < \lambda) + v_t \quad (17)$$

the two-stage least squares estimator for the threshold is found by solving

$$\hat{\lambda} = \min_{\lambda \in \Lambda} \hat{v}(\lambda)' \hat{v}(\lambda).$$

3. With the estimate  $\hat{\lambda}$ , GMM is then applied to each subsample using (16).

Denoting the smallest and largest inflation thresholds respectively by  $\underline{\lambda}$  and  $\bar{\lambda}$ ,  $\Lambda$  is then defined, as in TD, as  $\Lambda = [\underline{\lambda}, \bar{\lambda}]^3$ . The three panels of Figure 1 depict this situation. In the upper panel we have the interest rate which will be classified into two regimes according to the estimated threshold. The search over  $\Lambda$  will be performed according to the inflation rate variable which appears in the middle panel (ignoring for now the two horizontal lines). As can be seen, the lowest inflation rate occurs early in 2002 and the highest inflation rate in late 1992. The lower panel (ignoring for now the two vertical lines) tracks the value of the sum of squared residuals (see step 2 just above) over this search. We can see in the lower panel that at the start and end of the sample, or for low and high values of the inflation rate, the criterion function is not well behaved. This is in part due to the limited number of observations available in each regime.

A trimming rule frequently used in both Monte Carlo experiments and in empirical applications, is to drop a fraction of the smallest and largest values of the threshold. Typically a 10% or 15% rule is

---

as part A but with the default options, and hence only the first step of the two-step estimation procedure is performed. That is, the IV criterion function (11) is optimized without using the optimal weighting matrix. The latter is the approach actually taken by TD. The required two-step GMM-IV computations in Ox (lines 1 to 10), yield identical results with those of part A of the RATS code. It is important to note that only part A of the RATS code yields the proper standard errors of the estimates (as in line 10 of the Ox code).

<sup>3</sup>The lower and upper thresholds are chosen such that we have enough observations for estimation in each regime.



used. The estimated threshold can then be obtained by minimizing the SSR as in step 2 above over the relevant range of threshold

$$\hat{\lambda} = \min_{\lambda \in \Lambda^*} \hat{v}(\lambda)' \hat{v}(\lambda)$$

where  $\Lambda^*$  is the trimmed threshold variable.

This is shown in the middle and lower panels. The two horizontal lines represent this trimming rule while the corresponding vertical lines in the lower panel show the relevant values of the criterion function. As is clear from these pictures, trimming eliminates odd behaviour of the objective function (and of test statistics based on it).

This procedure, which gives consistent estimators, differs from the one used by TD in that their threshold is estimated directly by the minimization of a criterion function, which is the IV criterion function (11) and not the efficient GMM function (12) or (16).

## 4 Summary of the results

### 4.1 General discussion and specification tests

A summary of the estimation results is provided in Table 1 (standard errors are in parentheses). Estimation was performed, for the new models M4 and M2, allowing for heteroskedasticity and serial correlation using (16) with the automatic bandwidth selection procedure of Newey and West (1987) and Andrews and Monahan (1992)<sup>4</sup>. We used the same set of instruments as in TD (see above) but we also tried the instruments proposed by Clarida, Galí, and Gertler (1998) and Nelson (2001) with no significant changes in the results. The number of lags in the interest rate smoothing function was set to 1 in TD but we found it necessary to use 2 lags to remove any serial correlation in the error term of this process<sup>5</sup>. For the nonlinear model M2 we use the sequential estimation procedure described in the previous section with a trimming rule of 10%<sup>6</sup>.

Previous work by Clarida, Galí and Gertler (1998) for the UK during the period 1979 to the end of 1990 indicated that the interest rate setting rule was based on two factors, expected inflation and the German short rate. Nelson (2001) found that for the period 1992 to 1997, which is basically half of our sample, the policy rule followed by the Bank of England was one with an inflation coefficient greater than one and an output coefficient positive, providing support for a regime of inflation targeting. The

<sup>4</sup>We also allowed for only heteroskedasticity using (13) and the results were qualitatively similar.

<sup>5</sup>A key assumption in Caner and Hansen (2004) is that of stationarity. For this sample the null hypothesis of nonstationarity was rejected using the ADF test for inflation, the interest rate and the output gap. More on this later.

<sup>6</sup>A trimming rule of 15% was also used giving qualitatively similar results while a 5% rule gave regimes with too few observations to deem them reliable.

period investigated here, 1992:10 to 2003:1, saw important changes in policy structure including inflation targeting in 1992 and the independence of the Bank of England in 1997. Further, fluctuations (see Figure 1) in the interest rate and inflation were relatively modest (as compared to earlier periods). This relative stability might shed light on the likelihood of a multi-regime model.

Before discussing the results in details we performed a Durbin-Wu-Hausman test of the null hypothesis that the error terms are uncorrelated with all the regressors against the alternative that they are correlated with some of them but not correlated with the instruments. In our context, one might think that the forecast horizon of 3 months,  $k$ , is too short for monetary policy to have any influence on the economy (Gamber, Sinclair, Stekler, and Reid (2008)) and hence making IV unnecessary. Under the null hypothesis the OLS and IV estimators are consistent but under the alternative only IV estimators will be consistent. The DWH test statistic was 1858 with an asymptotic p value of 0. The null hypothesis is therefore rejected and hence either the inflation forecast can be treated as endogenous or some linear combination of the instruments have some explanatory power over and above the power of the regressors alone. Given this result we opt for the use of IV estimation. Although not reported here, the OLS estimates were different than those of the IV procedure, providing more evidence towards the non-equivalence of OLS and IV.

The columns labeled M4 and M4-new in Table 1 contain the results for the linear model of TD and ours, respectively. We used the same instruments, lag structure (1 lag) and time horizon for expected inflation (3 months) but estimation utilized efficient GMM. For this baseline model, we note first that the estimates are more precise than those obtained using the inefficient procedure (compare the standard errors of model M4 and M4-new) and that the Hansen-Sargan test for the validity of the extra instruments, or overidentifying restrictions, has a  $p$  value close to 1. Here we are testing the null hypothesis regarding the validity of the instruments. Note that a large  $p$  value implies that some or all of the instruments are not endogenous and therefore can be used in estimation. Our results indicate that, given the current information set (which includes the selected instruments), we can find the parameters  $[\alpha_0, \alpha_1, \alpha_2, \alpha_3]$  (and also  $[\delta, \gamma]$ ) that satisfy the Taylor rule.

In terms of the key underlying parameters, the estimation results for models M4 and M4-new are perhaps counterintuitive. The inflation parameter is below unity (and not significant) indicating that the Bank of England has, at best, accommodated inflation. The output gap parameter however is positive and significant implying a policy rule that is stabilizing with respect to the output gap. The estimate of the smoothing parameter (high  $\alpha_1$ ) indicates that interest rate deviations from target linger, perhaps lending support to the view that the central bank resorts to interest rate smoothing. The estimate of the intercept of the Taylor rule, in model M4-new, is 4.72 (standard error is 0.57) implying a long run equilibrium real interest rate of 2.89% which is on the low side (see Figure 2).

Model M4 was estimated again with two changes. The results appear in column M4-new-2<sup>7</sup>. First, two lags in the smoothing process were required so that the error term of that equation was uncorrelated. Second, a search over the time horizon for inflation,  $k$ , and for the output gap,  $q$ , was performed (for values of  $k$  and  $q$  ranging from 1 to 12). Candidate models were chosen using criteria such as the  $\bar{R}^2$  and the standard error of the regression (see also Orphanides (2001)). The choice of  $q$  was not influential and so was left unchanged. The best-fitting horizon for expected inflation was found to be 9 months, which is in line with the 12-month horizon used by Clarida, Galí, and Gertler (1998) who used monthly data and the 3-quarter horizon used by Nelson (2001) with a quarterly frequency<sup>8</sup>. The results appear under column M4-new-2. There is no evidence against the overidentifying restrictions implied by this model and, in addition, it yields estimates that have interesting economic interpretation. The inflation parameter, for example, is slightly above unity while the output parameter is significantly greater than 0. Nelson (2001) found similar results for the 1992-1997 period. Using data for the US, Inoue and Shintani (2006) using a bootstrap approach also found that the Fed was combating inflation.

## 4.2 Testing for nonlinearity

To test the linearity hypothesis, we compare the estimates from the two inflation regimes. If they are statistically equal, the null hypothesis is not rejected. Instead, TD compare the fit of the linear and nonlinear models using a sub-optimal, one-step, objective function. More precisely, our null hypothesis can be stated as

$$H_0 : \alpha = \beta \tag{18}$$

where  $\alpha$  and  $\beta$  can now be interpreted as vectors (recall expression (5)). Under the null the threshold effect in the Taylor rule vanishes. If the null hypothesis is rejected, we obtain the nonlinear model M2. We follow the approach of Caner and Hansen (2004), which is similar to the fixed bootstrap of Hansen (1996). This approach is summarized as follows:

1. For each  $\lambda \in \Lambda^*$  compute the estimates for the lower regime<sup>9</sup> and (regime 1) and the upper regime (regime 2) as well as their covariance matrices<sup>10</sup> ( $\hat{V}_1$  and  $\hat{V}_2$ ) and then form the Wald test statistic

$$W(\lambda) = (\hat{\alpha} - \hat{\beta})'(\hat{V}_1 + \hat{V}_2)^{-1}(\hat{\alpha} - \hat{\beta}). \tag{19}$$

---

<sup>7</sup>The coefficient on  $\alpha'_1$  is for the second autoregressive parameter.

<sup>8</sup>Results for the choice of  $k$  are available upon requests.

<sup>9</sup>In this paper the lowest and highest 10% of the observations are dropped.

<sup>10</sup>It should be noted that if one were to search for an optimal threshold, section 3.2, the resulting estimate would not be precise. This is due, in part, to the lack of dispersion in the threshold variable. The model was also estimated over the sample period 1992-2007 which yielded much more precise threshold estimate.

The supremum test is then

$$\sup W = \sup_{\lambda \in \Lambda^*} W(\lambda). \quad (20)$$

For each  $\lambda$  store the stacked residuals  $\hat{u}(\lambda)$ .

2. The p value is then obtained by simulation (see Hansen (1996)) in which the regressors are treated as fixed. A simulated dependent variable, one for each  $\lambda$  is created as  $y_t^*(\lambda) = \hat{u}(\lambda)\eta_t$  with  $\eta \sim N(0, 1)$ . Estimation then proceeds as before but with  $y_t^*$  replacing  $i_t$  and a simulated supremum test is computed

$$\sup W^* = \sup_{\lambda \in \Lambda^*} W(\lambda)^* \quad (21)$$

3. This is repeated 10,000 times and the p value is then computed as the fraction of the  $\sup W^*$  greater than  $\sup W$ .

This procedure yielded a  $\sup W = 2340$  with a simulated p value of 0.2759 indicating that we can reject the null hypothesis of linearity for significance levels greater than 27.59%<sup>11</sup>. Given this result, the null hypothesis is not rejected at any conventional significance levels. For the 1992-2003 time period, the data does not support the presence of a nonlinear Taylor rule.

For completeness we also report in Table 1 the estimates for the nonlinear models in columns labeled M2 (TD's results), M2-new (TD's specification with the new estimation procedure) and M2-new-2 (our specification with 2 lags for smoothing and a forecast horizon of 9 months and the new estimation procedure). The results in the upper panel are those for the  $\pi_{t-1} \geq \lambda$  regime while those for the lower panel are for the lower regime ( $\pi_{t-1} < \lambda$ ). The inflation threshold for all 3 models was 3.1%. We note that our parameter estimates in the last two columns are estimated more precisely and that the underlying inflation and output parameters are plausible (consider for example the last column M2-new-2). Further, the move from the inefficient procedure to the efficient one (moving from column M2 to M2-new for example) provides non trivial changes in the magnitudes of some parameters.

### 4.3 Testing for nonstationarity and nonlinearity

A nonstationary interest rate can be accommodated if we allow a unit root in the interest rate process. That is, we set  $\alpha_1 = 1$  in the single regime model or  $\alpha_1 = 1$  and  $\beta_1 = 1$  in the two regimes model. Under a unit root, the interest rate would behave like a random walk. Given that one of TD's preferred specification contained a two regime model with a unit root in the lower regime we will use the approach of Caner and Hansen (2001) that considers threshold autoregressions with a unit root. They look at a model just like (5) but allow for nonstationarity in either or both regimes. Estimation was performed using least

---

<sup>11</sup>These results are for the model M4-new-2. For model M4-new the Wald test is 2278 with a p value of 0.1612.

squares and so we extend their setting to the context of estimation using instrumental variables. Caner and Hansen (2001) consider the null of linearity  $\alpha = \beta$ , allow for a unit root and develop a Wald test similar to (21) that we present as an additional diagnostic tool. Their bootstrap procedure is different than the fixed regressors bootstrap explained above. More specifically, the bootstrap procedure would be as follows:

B1. Compute  $\sup W$  as described above.

B2. Estimate (4) and obtain  $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3$  and  $\tilde{u}_t$ . Generate the bootstrap data as

$$i_t^* = \tilde{\alpha}_0 + \tilde{\alpha}_1 i_{t-1}^* + \tilde{\alpha}_2 \pi_{t,k} + \alpha_3 x_{t,q} + u_t^*$$

where the  $u_t^*$  are random draws from the distribution of  $\tilde{u}_t$  and initial values can be set to available pre-sample values.

B3. Compute  $\sup W^*$ , repeat (10,000 times) and calculate the p value.

If the interest rate is stationary, this unconstrained bootstrap will achieve the correct first-order asymptotic distribution. If the interest rate is nonstationary they suggest to impose the unit root  $\alpha_1 = 1$  in step B2 above to obtain the constrained bootstrap.

Because we have two endogenous variables in our model, the bootstrap data generating process must be such that it generates bootstrap realizations of the main endogenous variable,  $i_t$ , and of the endogenous explanatory variable  $\pi_{t,k}$ . For this we follow the procedure of Davidson and MacKinnon (2003). Let the main dependent variable  $i_t$  be denoted by  $y$ , the endogenous explanatory variable  $\pi_{t,k}$  be denoted by  $Y$ ,  $W$  be a matrix containing the instruments and  $X$  a matrix containing the key regressors (here a constant, the output gap, the lag of the interest rate and  $\pi_{t,k}$ ). The reduced form equation for  $Y$  is then

$$Y = W\Pi_2 + V_2$$

where  $W$  is  $n \times l$ ,  $\Pi_2$  is  $l \times 1$  and the error term  $V_2$  is  $n \times 1$ . The nonparametric bootstrap data generating process would consist of

1. Estimate  $y = X\theta + u$  and obtain  $\tilde{u}$  and  $\tilde{\theta}$ .
2. Estimate  $Y = W\Pi_2 + V_2$  and obtain  $\tilde{\Pi}_2$  and  $\tilde{V}_2$ . Let  $\tilde{V} = [\tilde{u} \quad \tilde{V}_2]$ .
3. For each bootstrap replication, draw complete rows,  $\tilde{V}_t^*$ , at random from  $\tilde{V}$ .
4. Generate the endogenous explanatory variable as  $Y^* = W\tilde{\Pi}_2 + V_2^*$ .
5. Generate the main dependent variable as  $y^* = X^*\tilde{\theta} + u^*$  where the  $X^*$  indicates that some regressors (here the inflation rate) may be endogenous and come from step 4.

6. Because some of the non-endogenous explanatory variables are lagged dependent variables and lags of the endogenous explanatory variables, the bootstrap samples are generated recursively.

Steps 1 to 5 would then replace step B2 above. The analysis of the interest rate as possibly a nonlinear and/or nonstationary time series reveals first that for the Wald test of a threshold effect,  $\sup W = 2278$ , we obtain an unconstrained bootstrap p value of 0.9681 and a constrained bootstrap p value of 1. The null hypothesis of a linear model is again not rejected, taking into account the possibility of a unit root.

We can go a step further and also test for a unit root while allowing for a threshold. Caner and Hansen (2001) however showed in their simulation study that, given no threshold (as we have here), the tests for a unit root have no useful power and therefore very seldom reject the unit root hypothesis. We conclude, therefore, that we do not reject the linearity hypothesis whether we assume the presence of a unit root or not.

## 5 Conclusion

In this paper we considered a nonlinear forward-looking Taylor rule. This required estimation via an efficient instrumental variables method that allowed for nonlinearity. For this we followed the approach of Caner and Hansen (2004) which deals precisely with instrumental variables estimation of threshold models. Their estimation strategy was extended in a time series context. We took a second look at the novel work of Taylor and Davradakis (2006) that allowed central banks to have different monetary reaction functions in high and low inflation regimes and found that the null hypothesis of a linear model was not rejected for the sample period considered in Taylor and Davradakis (2006). This suggests that, if the Bank of England had been following a Taylor rule, the latter was linear. The underlying key parameters indicate that the Bank was concerned with the output gap, as well as with inflation during the 1992-2003 period. This is in accord with the inflation-targeting regime that started in 1992.

Interesting avenues of research include extending the nonlinear framework for the Taylor rule to real-time data. Orphanides (2001), Orphanides (2003) and more recently Molodtsova, Nikolsko-Rzevskyy, and Papell (2008) and Nikolsko-Rzevskyy (2009) considered linear Taylor rules with real-time data. In all cases estimation is simplified as the use of real-time data allows the researchers to use a nonlinear least squares procedure as opposed to some IV or GMM procedure. A second possibility, deals with identification issues in forward-looking Taylor rules. An and Schorfheide (2007) argued that there are important parameter identification issues in DSGE models. Given that monetary policy rules characterize many DSGE models, identification might be a problem. Mavroeidis (2008) used identification-robust procedures in his evaluation of the US Taylor rule and found that identification issues should be taken seriously. Inoue

and Rossi (2008) proposed a test for weak identification in nonlinear models and apply it also to the US Taylor rule. They found that identification issues were not a problem for the GMM parameters (our  $\alpha$ 's and  $\beta$ 's) but an issue for the underlying key parameters (our inflation and output parameters). We implemented the first-stage  $F$  statistic aimed at detecting weak identification and obtained a value of 7.26. The rule-of-thumb for this test is that a value less than 10 is an indication that the instruments are weak and the parameters might not be well identified. Given this result, the use of real-time data seems important. This will be the focus of our future research.

## 6 Appendix

Estimation of Model M4 in Regression Analysis of Time Series (RATS):

```
* Part A: Two-step procedure with nlsystem

disp 'First-step GMM-IV estimates'
linreg(inst) r start end resid
# constant r{1} inf{-3} gap
mcov(lags=2,lwindow=neweywest) / resid
# constant r{1 to 6 9 12} ppi{1 to 6 9 12} inf{1 to 6 9 12} gap{1 to 6 9 12} RPI{1 to 6 9 12} RPIY{1 to 6 9 12}

nonlin e0 e1 e2 e3
com e0=e1=e2=e3=0
frml trule r = e0 + e1*r{1} + e2*inf{-3} + e3*gap
nlsystem(instruments,zudep,secondstep,swmatrix=inv(%cmom)) start end trule

* Part B: One-step procedure with nlsystem

nonlin e0 e1 e2 e3
com e0=e1=e2=e3=0
frml trule r = e0 + e1*r{1} + e2*inf{-3} + e3*gap
nlsystem(instruments) start end trule
```

Estimation of Model M4 in Ox:

```
print("First-step GMM-IV estimates");

1 WW=W'*W;

2 WX=W'*X;

3 WY=W'*Y;

4 iWW=invertsym(WW);

5 beta=invertsym(WX'*iWW*WX)*WX'*iWW*WY;

6 print("Second-step GMM-IV estimates");

7 Sigma=nw(beta); // Newey-West Estimator

8 iSigma=invertsym(Sigma);

9 beta=invertsym(WX'*iSigma*WX)*WX'*iSigma*WY;

10 cov=T*invertsym(WX'*iSigma*WX); // Variance/Covariance Matrix Estimation
```



Figure 1: Threshold estimation and the trimming requirement

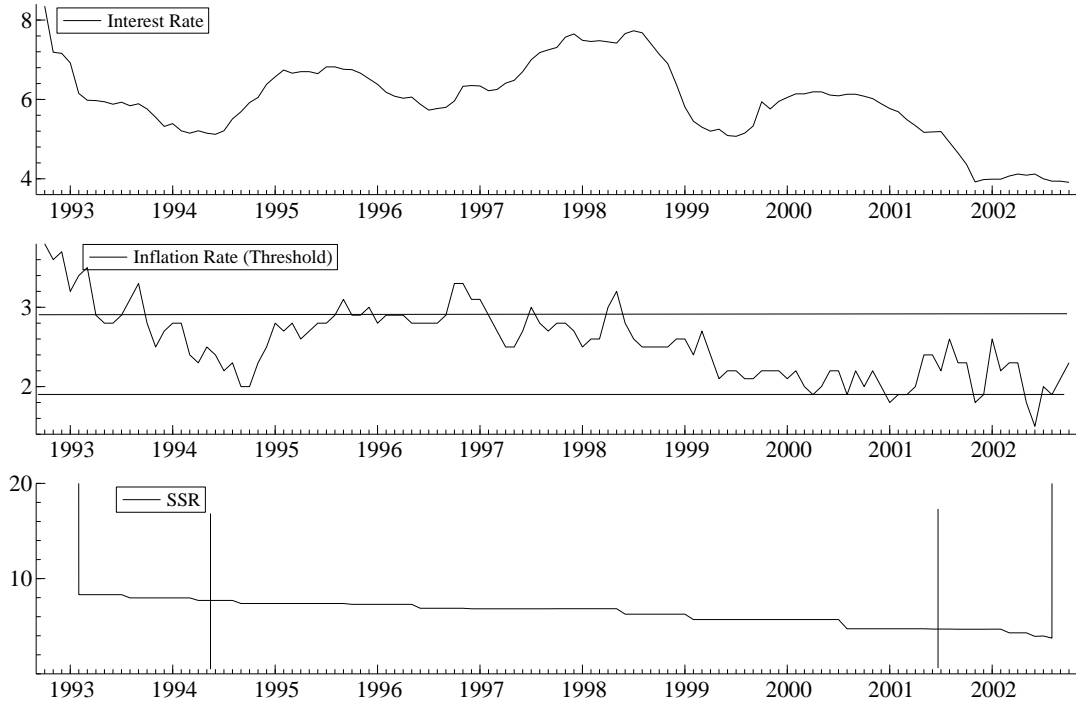


Figure 2: Real Interest Rate

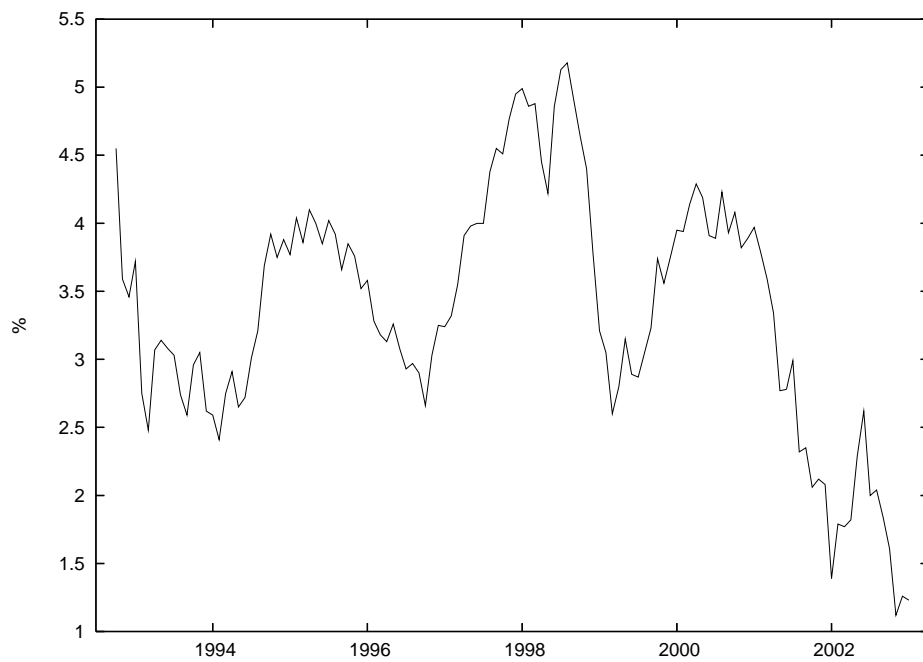


Table 1: Estimation Results

	M4	M4-new	M4-new-2	M2	M2-new	M2-new-2
$\alpha_0$	0.4361 (0.1661)	0.4298 (0.0445)	0.3238 (0.0472)	1.6079 (0.6191)	2.6813 (0.2198)	0.6229 (0.2099)
$\alpha_1$	0.9087 (0.0263)	0.9090 (0.0103)	1.3675 (0.0217)	0.4402 (0.0724)	0.5579 (0.0209)	0.8133 (0.0387)
$\alpha'_1$			-0.4667 (0.0192)			-0.2843 (0.0355)
$\alpha_2$	0.0227 (0.0773)	0.0243 (0.0230)	0.1021 (0.0162)	0.7029 (0.2988)	0.0449 (0.0707)	0.8827 (0.0895)
$\alpha_3$	0.0832 (0.0257)	0.0828 (0.0091)	0.0195 (0.0042)	0.2432 (0.1010)	0.0916 (0.0156)	0.0149 (0.0304)
Inflation Parameter	0.2494 (0.8071)	0.2672 (0.2312)	1.0295 (0.1477)	1.2558 (0.4360)	0.1016 (0.1594)	1.8742 (0.1591)
Output Parameter	0.9120 (0.3345)	0.9105 (0.1258)	0.1963 (0.0450)	0.4345 (0.1676)	0.2072 (0.0351)	0.0316 (0.0653)
$\beta_0$				-0.1523 (0.1469)	-0.0879 (0.0371)	-0.0313 (0.0365)
$\beta_1$				0.9928 (0.0226)	0.9811 (0.0067)	1.4839 (0.0255)
$\beta'_1$						-0.5209 (0.0265)
$\beta_2$				0.0686 (0.0621)	0.0684 (0.0149)	0.0978 (0.0107)
$\beta_3$				0.0547 (0.0203)	0.0482 (0.0061)	0.0141 (0.0034)
Inflation Parameter				9.6266 (0.0968)	3.6289 (1.0501)	2.6442 (0.4175)
Output Parameter				7.6725 (0.0359)	2.5548 (0.8434)	0.3815 (0.1018)
Hansen-Sargan Test	72.6781 (0.0056)	13.8272 (0.9999)	14.5369 (0.9999)			

## References

- An, S. and F. Schorfheide (2007): “Bayesian analysis of DSGE models,” *Econometric Reviews*, 26, 113–172.
- Andrews, D. and J. Monahan (1992): “An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator,” *Econometrica*, 60, 953–966.
- Andros, K., T. Stengos, and C. M. Tan (2009): “Structural threshold regression,” Working paper.
- Bunzel, H. and W. Enders (2005): “Is the Taylor rule missing? a statistical investigation,” Working Paper 05015, Iowa State University.
- Caner, M. and B. Hansen (2001): “Threshold autoregression with a unit root,” *Econometrica*, 69, 1555–1596.
- Caner, M. and B. Hansen (2004): “Instrumental variable estimation of a threshold model,” *Econometric Theory*, 20, 813–843.
- Chan, K. (1993): “Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model,” *The Annals of Statistics*, 21, 520–533.
- Clarida, R., J. Galí, and M. Gertler (1998): “Monetary policy rules in practice: Some international evidence,” *European Economic Review*, 42, 1033–1067.
- Clarida, R., J. Galí, and M. Gertler (2000): “Monetary policy rules and macroeconomic stability: Evidence and some theory,” *Quarterly Journal of Economics*, 115, 147–180.
- Davidson, R. and J. MacKinnon (2003): *Econometric Theory and Methods*, New York: Oxford University Press.
- Doornik, J. (2007): *Object-Oriented Matrix Programming Using Ox*, London: Timberlake Consultants Press and Oxford, 3rd edition.
- Estima (2006): *RATS Version 6.2 User’s Guide*, Evanston.
- Gamber, E., T. Sinclair, H. Stekler, and E. Reid (2008): “Multivariate forecast errors and the Taylor rule,” Working Paper 2008-002, George Washington University.
- Hansen, B. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” *Econometrica*, 64, 413–430.
- Hansen, B. (2000): “Sample splitting and threshold estimation,” *Econometrica*, 68, 575–603.

- Hansen, L., J. Heaton, and A. Yaron (1996): “Finite-sample properties of some alternative GMM estimators,” *Journal of Business and Economic Statistics*, 14, 262–280.
- Inoue, A. and B. Rossi (2008): “Testing for weak identification in possibly nonlinear models,” Working paper.
- Inoue, A. and M. Shintani (2006): “Bootstrapping GMM estimators for time series,” *Journal of Econometrics*, 133, 531–555.
- Mavroeidis, S. (2008): “Monetary policy rules and macroeconomic stability: some new evidence,” Working paper, Brown University.
- Molodtsova, T., A. Nikolsko-Rzevskyy, and D. Papell (2008): “Taylor rules with real-time data: A tale of two countries and one exchange rate,” *Journal of Monetary Economics*, 55, 563–579.
- Nelson, E. (2001): “UK monetary policy 1972-1997: A guide using Taylor rules,” Working Paper 2931, CEPR Discussion Paper.
- Newey, W. and K. West (1987): “A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708.
- Nikolsko-Rzevskyy, A. (2009): “Monetary policy evaluation in real time: Forward-looking Taylor rules without forward-looking data,” Working paper, University of Memphis.
- Orphanides, A. (2001): “Monetary policy rules based on real-time data,” *American Economic Review*, 91, 964–985.
- Orphanides, A. (2003): “Historical monetary policy analysis and the Taylor rule,” *Journal of Monetary Economics*, 50, 983–1022.
- Sargan, J. (1958): “The estimation of economic relationships using instrumental variables,” *Econometrica*, 26, 393–415.
- Taylor, J. (1993): “Discretion versus policy rules in practice,” *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.
- Taylor, J. (1999): “The robustness and efficiency of monetary policy rules as guidelines for interest rate setting by the European central bank,” *Journal of Monetary Economics*, 43, 655–679.
- Taylor, M. and E. Davradakis (2006): “Interest rate setting and inflation targeting: Evidence of a nonlinear Taylor rule for the United Kingdom,” *Studies in Nonlinear Dynamics and Econometrics*, 10, 1–18.