

Saikkonen, Pentti; Ripatti, Antti

Working Paper

On the estimation of Euler equations in the presence of a potential regime shift

Bank of Finland Discussion Papers, No. 6/1999

Provided in Cooperation with:

Bank of Finland, Helsinki

Suggested Citation: Saikkonen, Pentti; Ripatti, Antti (1999) : On the estimation of Euler equations in the presence of a potential regime shift, Bank of Finland Discussion Papers, No. 6/1999, ISBN 951-686-613-1, Bank of Finland, Helsinki,
<https://nbn-resolving.de/urn:nbn:fi:bof-20140807499>

This Version is available at:

<https://hdl.handle.net/10419/211837>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

BANK OF FINLAND DISCUSSION PAPERS

6/99

Pentti Saikkonen – Antti Ripatti

Research Department
3.6.1999

On the Estimation of Euler Equations
in the Presence of a Potential Regime Shift

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

Pentti Saikkonen* – Antti Ripatti

Research Department
3.6.1999

On the Estimation of Euler Equations in the Presence of a Potential Regime Shift

The views expressed are those of the authors and do not necessarily correspond to the views of the Bank of Finland

* University of Helsinki

ISBN 951-686-613-1
ISSN 0785-3572

Suomen Pankin monistuskeskus
Helsinki 1999

On the Estimation of Euler Equations in the Presence of a Potential Regime Shift

Bank of Finland Discussion Papers 6/99

Pentti Saikkonen
University of Helsinki

Antti Ripatti
Research Department

Abstract

The concept of a peso problem is formalized in terms of a linear Euler equation and a nonlinear marginal model describing the dynamics of the exogenous driving process. It is shown that, using a threshold autoregressive model as a marginal model, it is possible to produce time-varying peso premia. A Monte Carlo method and a method based on the numerical solution of integral equations are considered as tools for computing conditional future expectations in the marginal model. A Monte Carlo study illustrates the poor performance of the generalized method of moment (GMM) estimator in small and even relatively large samples. The poor performance is particularly acute in the presence of a peso problem but is also serious in the simple linear case.

Keywords: peso problem, Euler equations, GMM, threshold autoregressive models

Eulerin yhtälöiden estimoinnista mahdollisten regiimin muutosten tapauksessa

Suomen Pankin keskustelualoitteita 6/99

Pentti Saikkonen
Helsingin yliopisto

Antti Ripatti
Tutkimusosasto

Tiivistelmä

Tutkimuksessa muotoillaan peso-ongelma lineaarisen Eulerin yhtälön ja epälineaarisen, eksogeenista ohjausprosessia kuvaavan marginaalimallin avulla. Ajassa vaihteleva pesopreemio syntyy mm. silloin, kun marginaalimalli on muodoltaan autoregressiivinen kynnysmalli. Tutkimuksessa tarkastellaan lisäksi Monte Carlo -menetelmää ja integraaliyhtälöiden numeeriseen ratkaisuun perustuvaa menetelmää laskea marginaalimallin ennusteet. Tutkimuksessa muotoiltuun malliin nojaavat Monte Carlo -kokeet viittaavat siihen, että yleistettyyn momenttimenetelmään, GMM:ään, perustuvat estimaattorit toimivat huonosti pienissä otoksissa. Tämä pätee paitsi peso-ongelmaisiin malleihin myös tavallisiin lineaarisiin malleihin.

Asiasanat: peso-ongelma, Eulerin yhtälö, GMM, autoregressiiviset kynnysmallit

Contents

Abstract.....	3
1 Introduction.....	7
2 Model	8
3 Computation of conditional expectations	13
4 ML estimation.....	16
5 GMM estimation.....	18
6 Simulation results	22
7 Conclusions.....	32
Appendix	33
References	34

1 Introduction

Euler equations appear in a variety of modern dynamic economic models developed in areas such as consumption, investment and inflation to name a few. Well-known examples are Hall's (1978) life cycle -permanent income model, Abel's (1982) model of investment with adjustment costs and Cagan's (1956) hyperinflation model. In a simple special case an Euler equation states that the current value of a variable of interest is a linear function of its future rational expected value and the current value of a so called forcing variable. The econometric analysis of an Euler equation often suffers from a problem commonly referred to as a peso problem. This means a situation where the potential of a regime shift in the forcing variable affects agents' expectations. From a statistical point of view the probability of such a regime shift has to be small because it is assumed that no regime shift has actually occurred during the sample period. However, if the potential of a regime shift has still affected agents' expectations ignoring this in the econometric analysis means using a misspecified model and, consequently, inefficient or even invalid econometric methods. This in turn can lead to seriously misleading conclusions.

This paper is concerned with some econometric aspects related to Euler equations in the presence of a peso problem. There are two conceivable extensions of conventional linear time series models which could be entertained to model the forcing variable and thereby expectations based on it in this situation. The first one is an autoregressive model in which the mean value changes between two regimes according to a Markov chain. This model, discussed by Hamilton (1993, 1994, chapter 22), has recently been employed in various contexts but, at least in its simplest form, turns out to be unsuitable for our purposes. As a second model we shall therefore consider a threshold autoregressive model in which the intercept term can switch between two regimes. This model seems to be capable of allowing for the peso problem and will therefore be explored in some detail.

The plan of the paper is as follows. Section 2 introduces the problem and discusses the above mentioned models. In order to make the main points easy to follow several simplifying assumptions will be made. Section 3 deals with the computation and properties of conditional expectations in the case where a threshold autoregressive model is assumed for the forcing variable. These results are employed in Section 4 where Gaussian maximum likelihood estimation of the parameters in both the Euler equation and the threshold autoregressive model of the forcing variable is discussed. The same set-up is used in Section 5 to study the generalized method of moments (GMM) estimation of the parameters in the Euler equation. Section 6 studies small sample properties of various GMM estimators by Monte Carlo simulation. Conclusions are given in Section 7.

2 Model

Consider the two variables y_t and x_t related by the simple Euler equation

$$y_t = v + \alpha E_t^* y_{t+1} + \beta x_t \quad (2.1)$$

where E_t^* denotes the conditional expectation with respect to a information set available to economic agents at time t . Assuming that $0 < \alpha < 1$ equation (2.1) can be solved as

$$y_t = m + \beta \sum_{j=0}^{\infty} \alpha^j E_t^* x_{t+j} \quad (2.2)$$

where $m = v/(1-\alpha)$ and $E_t^* x_t = x_t$. For econometric analysis we often have to replace the information set used to define E_t^* by a smaller counterpart available to the econometrician. Since we assume that the two variables y_t and x_t are only observed this information set is here $\{y_s, x_s; s \leq t\}$. Instead of (2.2) we then have

$$y_t = m + \beta \sum_{j=0}^{\infty} \alpha^j E_t x_{t+j} + \varepsilon_t \quad (2.3)$$

where E_t is the conditional expectation with respect to the information set $\{y_t, x_t; s \leq t\}$ and

$$\varepsilon_t = \beta \sum_{j=1}^{\infty} \alpha^j E_t^* x_{t+j} - \beta \sum_{j=1}^{\infty} \alpha^j E_t x_{t+j}. \quad (2.4)$$

Clearly, ε_t has the martingale difference property

$$E_t \varepsilon_t = 0. \quad (2.5)$$

For simplicity, we also assume that there is no Granger causality from current and lagged values of y_t to x_t so that

$$E(x_{t+j} | y_{s+1}, x_s; s \leq t) = E(x_{t+j} | x_s; s \leq t), j \geq 1. \quad (2.6)$$

We are interested in modelling the effects of the peso problem so that the potential of a regime shift affects the agents' expectations about the forcing variable x_t although no regime shift has occurred during the sample period $t = 1, \dots, T$. It is fairly obvious that conventional linear models cannot be used to allow for this feature. A conceivable possibility might be to assume that the forcing variable x_t follows an autoregressive (AR) process whose mean value changes between two regimes according to a (homogeneous) Markov chain in such a way that the

probability of a shift to the higher level regime is very small (see eg Hamilton (1993) or (1994, chapter 22) for the this model). Specifically, one might specify

$$x_t = \mu_{s_t} + z_t \quad (2.7)$$

where z_t is a zero mean stationary AR process and s_t is a two-state Markov chain, independent of the z_t process. The transition probabilities of s_t are

$$P\{s_t = 1 \mid s_{t-1} = 1\} = p_{11}$$

$$P\{s_t = 2 \mid s_{t-1} = 1\} = p_{12}$$

$$P\{s_t = 1 \mid s_{t-1} = 2\} = p_{21}$$

$$P\{s_t = 2 \mid s_{t-1} = 2\} = p_{22}$$

where $s_t = 1$ indicates the regime from which the observations have been obtained and $s_t = 2$ indicates the higher level regime to which the process can potentially shift. Thus, $\mu_1 < \mu_2$ and the probabilities p_{11} and p_{12} are close to one and zero, respectively.

Even though the above model may look suitable for our purposes it does not work because we assume that no regime shift has occurred during the sample period and information of this is available. Indeed, under these assumptions the conditional expectations in (2.3) have to be calculated conditional on the known regime and, since we have assumed (2.6), the calculations in Hamilton (1993, p. 253) show that the infinite sum in (2.3) is the same as in the case where x_t follows a linear AR process except for an additive constant.¹ This constant is induced by the changing mean value μ_{s_t} in (2.7) and it describes the effect of the peso problem in this model. However, this effect cannot be estimated or taken into account because it cannot be separated from the intercept term m in (2.3). Thus, in this approach the effect of the peso problem is not identified. The situation might change if the homogeneity assumption of the Markov chain used to model the potential regime shift were abandoned and the transition probabilities were made dependent on observable variables. This would probably make the resulting model and its application rather complex so that another approach will be considered below.

We shall assume that x_t follows the threshold autoregressive (TAR) process

$$x_t = \mu + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \delta I(x_{t-1} \geq c) + \eta_t \quad (2.8)$$

where the indicator function $I(\cdot)$ takes the value one when the indicated statement is true and zero otherwise and η_t is a martingale difference sequence satisfying

$$E(\eta_t \mid x_s; s < t) = 0. \quad (2.9)$$

¹ Hamilton (1993) only considers a first order AR process but his treatment can easily be extended to any finite order AR process.

It is also assumed that the zeros of the polynomial $1 - \phi z - \dots - \phi_p z^p$ lie outside the unit circle. This assumption and some reasonably mild additional conditions on the error term η_t imply that x_t is geometrically ergodic and strong mixing with geometrically decaying mixing coefficients (see Masry and Tjøstheim (1995, Lemma 3.1)). For instance, assuming that η_t is Gaussian white noise is sufficient in this respect.

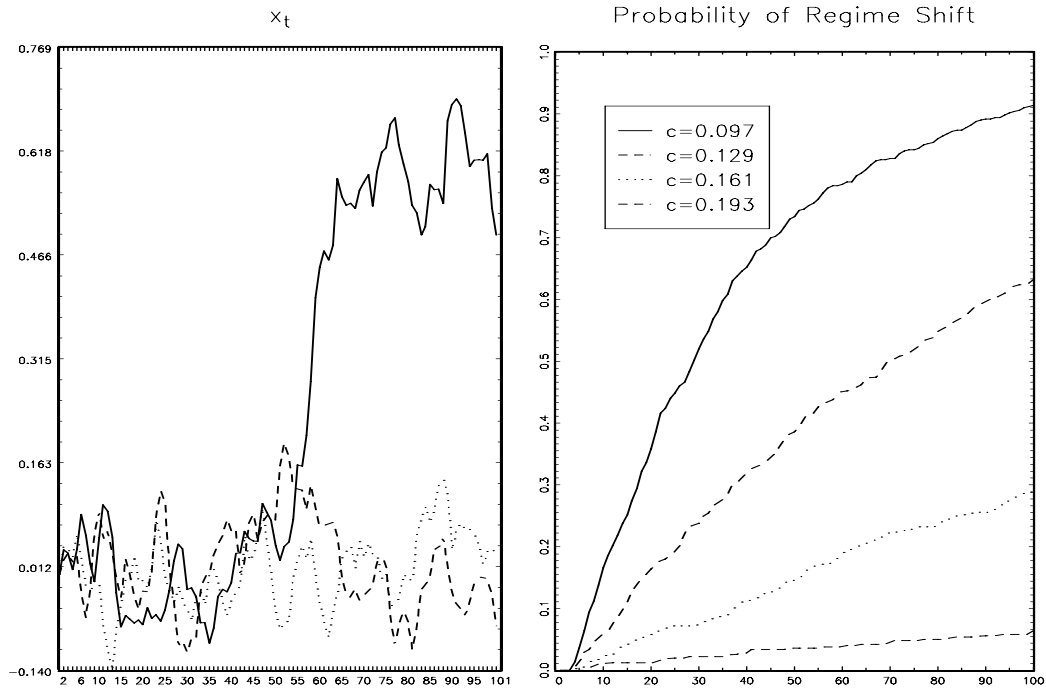
From (2.6), (2.8) and (2.9) one readily finds that $E(\eta_t | y_s, x_{s-1}; s \leq t) = 0$ which in conjunction with (2.3) yields $E\epsilon_t \eta_{t+j} = 0, j > 0$. On the other hand, since (2.5) and (2.8) give $E\epsilon_t \eta_{t-j} = 0, j \geq 0$, our assumptions imply that the error terms ϵ_t and η_s are uncorrelated for all t and s .

In the TAR process (2.8) possible regime shifts are modelled by the indicator function $I(x_{t-1} \geq c)$ and related parameter δ which is assumed to take on positive values. Our previous assumption that the observations of x_t all come from one regime means that the observed series x_1, \dots, x_T and also the initial values x_{-p+1}, \dots, x_0 are such that $x_t < c$ for all $t = -p+1, \dots, T$. Thus, within the sample the TAR model (2.8) actually reduces to a conventional linear AR model. However, at any point of time $x_{t-1} \geq c$ could have occurred and, since $\delta > 0$, a jump to a higher level regime could have occurred too. With suitable parameter values the probability of such a regime shift can be small and, consequently, it is possible to observe even fairly long realizations of x_t without a shift to the higher regime. This feature is illustrated in Figure 1 which shows realizations of the TAR process (2.8) and related probabilities of a regime shift. Even though the probability of a regime shift or the event $x_{t-1} \geq c$ is small it affects the conditional expectations $E_t x_{t+j}$ and thereby the behavior of y_t (see (2.3)). In particular, it is intuitively fairly obvious from (2.8) that when the value of x_{t-1} increases close to c the conditional expectations of future values of x_t also increase and, if the value of the parameter δ is large, the effect of this can be substantial and it differs from an additive constant. This fact can be seen in the simulated realizations of Figure 2 and it will also be seen formally in the next section.

If $\delta = 0$ a priori (2.8) reduces to a linear AR model and, as is well known, an analytic expression can be obtained for $E_t x_{t+j}$ and the infinite sum in (2.3). However, when $\delta > 0$ the situation is much more complicated and no analytic expressions are generally available. Computation of these conditional expectations is required, however, if maximum likelihood (ML) estimation of the parameters in (2.3) and (2.8), based on a chosen distributional assumption of the errors, is considered. The same is true if one wants to generate simulated data to see whether the idea used in the TAR model (2.8) appears reasonable. In the next section these computational issues are discussed in more detail. For this purpose, as well as for ML estimation, the martingale difference assumptions so far made of the error terms ϵ_t and η_t are not sufficient. Unless otherwise stated it will therefore be convenient to assume throughout the rest of this paper that $[\epsilon_t \ \eta_t]'$ is a sequence of independent and identically distributed (iid) random vectors. Sometimes a normal distribution will also be assumed to make the exposition more concrete.

Figure 1.

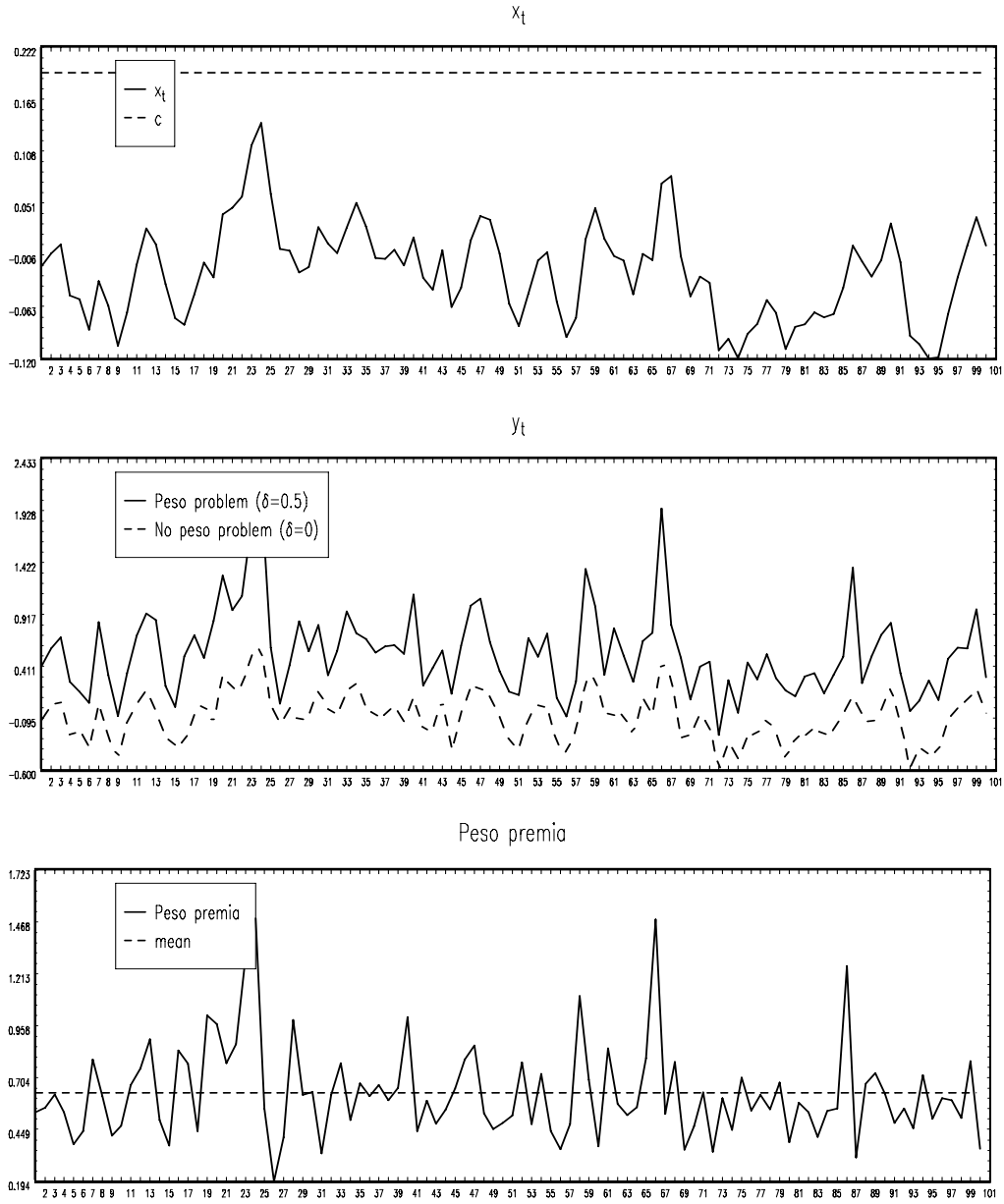
Threshold Autoregressive Model and the Probability of Regime Shift



The left panel shows three realization of the process x_t with $\eta_t \sim N(0, 0.001)$ and $T = 100$. The parameter values are $\nu = 0$, $\alpha = 1/1.05$, $\beta = 1$, $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$ (the roots are 0.7 and 0.4), $\delta = 0.25$ and $c = 0.193$. The right panel reports the probability of a regime shift before the period measured in horizontal axis. These probabilities are obtained by simulation and graphed for various choices of the threshold value.

Figure 2.

Realizations of the x_t and the y_t and the Time-Varying Peso Premia



The x_t process is simulated with $\eta_t \sim N(0, 0.001)$ and $T = 100$. The parameter values are $v = 0$, $\alpha = 1/1.05$, $\beta = 1$, $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$, $\delta = \{0, 0.5\}$ and $c = 0.193$. When computing the y_t process, the simulation method proposed by Clements and Smith (1997) is applied in the computation of $E_t x_{t+j}$ ($j = 1, \dots$).

3 Computation of conditional expectations

Consider the TAR model (2.8) with given parameter values. A simple way to compute the conditional expectations needed in (2.3) is to use the simulation method proposed by Clements and Smith (1997). In this approach $E_t x_{t+j}$ is replaced by a simulated counterpart for $j = 1, \dots, J$ and the value of J so large that the infinite sum can be truncated at $j = J$ with only a negligible error. The first step of this method is to simulate values of the error term η_t , which makes clear that the distribution of the errors has to be specified. Of course, for $j = 1$ it is not necessary to use simulation because

$$E_t x_{t+1} = \mu + \phi_1 x_t + \dots + \phi_p x_{t+1-p} + \delta I(x_t \geq c), \quad (3.1)$$

as is obvious from (2.8).

Even though the simulation method is simple to apply it is also reasonable to consider analytic methods because it is useful to understand analytic properties of the conditional expectations in (2.3). Here we assume that the initial values of x_t are taken from the stationary distribution and employ the companion form of (2.8). For simplicity, we also assume that $\mu = 0$ a priori. This means no loss of generality because in (2.8) we can use the transformations $x_t \rightarrow x_t - \mu_*$ and $c \rightarrow c - \mu_*$ where $\mu_* = \mu / (1 - \phi_1 - \dots - \phi_p)$. Thus, instead of (2.8) we can consider

$$\begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \vdots & & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{bmatrix} + \delta \begin{bmatrix} I(x_{t-1} \geq c) \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or

$$X_t = AX_{t-1} + \delta e_1 I(e_1' X_{t-1} \geq c) + N_t \quad (3.2a)$$

$$x_t = e_1' X_t \quad (3.2b)$$

where $e_1 = [1 \ 0 \ \dots \ 0]'$ ($px1$) and the rest of the notation is obvious. To simplify, write (3.2a) as

$$X_t = h(X_{t-1}) + N_t. \quad (3.3)$$

Thus, since η_t is a sequence of iid random variables, X_t is a Markov chain over \mathbb{R}^p . Denoting $E(X_{t+j}|X_t) = K_j(X_t)$ ($j \geq 1$) we can therefore conclude from the Chapman-Kolmogorov relation the recursion

$$K_j(X_t) = \int_{-\infty}^{\infty} K_{j-1}(z) p_{X_{t+1}|X_t}(z) dz, \quad j = 2, 3, \dots \quad (3.4)$$

where $p_{x_{t+1}|x_t}(z)$ signifies the conditional density function of X_{t+1} given X_t (see Tong, 1990, section 6.2.1).² The relation of this to the conditional expectations needed in (2.3) is seen from the identity $E_t x_{t+j} = e_1' K_j(X_t)$, $j \geq 1$. For $j = 2$ the right hand side of (3.4) depends on

$$K_1(z) = h(z) = Az + \delta e_1 I(e_1' z \geq c) \quad (3.5)$$

where the first equality is justified by (3.3) and the second one by (3.2a).

Equation (3.1) gives an expression of $E_t x_{t+j}$ for $j = 1$. Since we assume that $x_t < c$ the result is the same as in the case of a linear AR model. To see what happens when $j = 2$, note that from (3.4) and (3.5) it follows that

$$K_2(X_t) = Ah(X_t) + \delta e_1 \int_c^\infty g(z_1 - e_1' h(X_t)) dz_1 \quad (3.6)$$

where z_1 signifies the first component of the vector z and $g(\cdot)$ is the density function of η_t (cf. Tong, 1990, section 6.2.1). If we assume for concreteness that $\eta_t \sim N(0, \sigma_\eta^2)$ (3.6) yields

$$E_t x_{t+2} = e_1' Ah(X_t) + \delta \Phi\left(\frac{e_1' h(X_t) - c}{\sigma_\eta}\right) \quad (3.7)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The assumption $x_t < c$ implies that $h(X_t) = AX_t$ and the first term on the right hand side of (3.7) is the same as in the case where x_t follows a linear AR process. However, since $\delta > 0$ is assumed the second term on the right hand side of (3.7) is nonzero and gives an additional contribution to $E_t x_{t+2}$ even though $h(X_t) = AX_t$ holds, as in the linear case. In particular, when $e_1' h(X_t) = \phi_1 x_t + \dots + \phi_p x_{t-p+1}$ gets close to the value of the threshold parameter c the value of $E_t x_{t+2}$ gets larger than in the corresponding linear case. The larger the value of the parameter δ the stronger this effect is and it is clearly nonconstant. This provides a formal explanation for the similar intuitive conclusion given in the previous section.

Above we have seen how the nonlinearity of the TAR model (2.8) affects the conditional expectations $E_t x_{t+j}$ when $j = 1, 2$. This can be pursued further to see a similar effect on the infinite sum in (2.3). To this end, we first notice that the first term on the right hand side of (3.6) equals $AK_1(X_t)$. This suggests the recursion

$$K_j(X_t) = AK_{j-1}(X_t) + \delta e_1 k_j(X_t), \quad j = 2, 3, \dots \quad (3.8a)$$

$$k_j(X_t) = \int_{-\infty}^\infty k_{j-1}(z_1, x_t, \dots, x_{t-p+2}) g(z_1 - e_1' h(X_t)) dz_1 \quad (3.8b)$$

² Here, as well as later, the existence of the stated density function is assumed.

where $K_1(X_t)$ is obtained from (3.5) and $k_1(z) = I(e'_1 z \geq c)$ (cf. (3.6)). The validity of this recursion can be readily checked by induction (see the appendix). Equation (3.8a) can be solved to yield

$$K_j(X_t) = A^{j-1}K_1(X_t) + \delta \sum_{i=0}^{j-2} A^i e_1 k_{j-i}(X_t), \quad j \geq 2.$$

Since $E_t x_{t+j} = e'_1 K_j(X_t)$ and since $K_1(X_t) = AX_t$ is assumed it follows from this that

$$\begin{aligned} \sum_{j=0}^{\infty} \alpha^j E_t x_{t+j} &= e'_1 X_t + \alpha e'_1 A X_t + e'_1 \sum_{j=2}^{\infty} \alpha^j A^j X_t + \delta e'_1 \sum_{j=2}^{\infty} \alpha^j \sum_{i=0}^{j-2} A^i e_1 k_{j-i}(X_t) \\ &= e'_1 (I_p - \alpha A)^{-1} X_t + \delta e'_1 \sum_{i=0}^{\infty} \alpha^i A^i \sum_{j=2}^{\infty} \alpha^j e_1 k_j(X_t). \end{aligned}$$

Thus, we can write

$$\sum_{j=0}^{\infty} \alpha^j E_t x_{t+j} = e'_1 (I_p - \alpha A)^{-1} X_t + \delta e'_1 (I_p - \alpha A)^{-1} e_1 \sum_{j=2}^{\infty} k_j(X_t). \quad (3.9)$$

The latter term on the right hand side is due to the nonlinearity of the TAR model (2.8) and can be interpreted as a measure of the peso problem in model (2.3).

Arguments used to obtain (3.6) from (3.4) and (3.5) can also be used to obtain $e'_1 K_3(X_t) = E_t x_{t+3}$ and furthermore $e'_1 K_j(X_t) = E_t x_{t+j}$ for $j \geq 4$. Obviously, these results will get involved and cannot be given in closed form. However, even if the above derivations are not used for practical computation they can be used to study analytic properties of the infinite sum in model (2.3). This is needed to justify the application of conventional ML estimation theory, as will be discussed in the next section.

We close this section with a remark on the infinite sum on the right hand side of (3.9). Multiplying both sides of equation (3.8b) by α^j , summing over $j \geq 2$ and using the definition $k_1(z) = I(e'_1 z \geq c)$ we find that

$$\begin{aligned} \sum_{j=2}^{\infty} \alpha^j k_j(X_t) &= \alpha \int \sum_{j=2}^{\infty} k_j(z_1, x_t, \dots, x_{t-p+2}) g(z_1 - e'_1 h(X_t)) dz_1 \\ &\quad + \alpha^2 \int_c^{\infty} g(z_1 - e'_1 h(X_t)) dz_1. \end{aligned} \quad (3.10)$$

Thus, we have an integral equation for the infinite sum on the right hand side of (3.9). One could try to solve this integral equation directly and, as equation (3.9) shows, thereby be able to compute the infinite sum in model (2.3) without computing the individual summands $E_t x_{t+j}$. Finding out the feasibility of this approach requires expertise on numerical solution of integral equations, a topic outside the scope of this paper. For a comprehensive reference, see Baker (1977).

4 ML estimation

The discussion in the previous section shows that the infinite sum in (2.3) is a function of the vector X_{t-1} , the parameters of the x_t process in (2.8) and the parameter α . Thus, we can write

$$\sum_{j=0}^{\infty} \alpha^j E_t x_{t+j} = f^*(X_{t-1}; \alpha, \gamma, \delta, c, \sigma_\eta^2) \quad (4.1)$$

where $\sigma_\eta^2 = E\eta_t^2$ and the parameter vector $\gamma = [\mu, \phi_1, \dots, \phi_p]'$ contains the regression coefficients in (2.8) which can be estimated from this model even if $x_t < c$ is assumed. We can thus consider (2.3) as a nonlinear regression model

$$y_t = f(X_{t-1}; \psi, \gamma, \sigma_\eta^2) + \varepsilon_t, \quad t = 1, \dots, T \quad (4.2)$$

where $\psi = [m \ \beta \ \alpha \ \delta \ c]'$ and $f(X_{t-1}; \psi, \gamma, \sigma_\eta^2) = m + \beta f^*(X_{t-1}; \alpha, \gamma, \delta, c, \sigma_\eta^2)$. For simplicity, we shall use the notation $f(X_{t-1}; \cdot) = f_{t-1}(\cdot)$ and similarly for $f^*(X_{t-1}; \cdot)$ when there is no need to be explicit about the vector X_{t-1} .

For ML estimation we have to supplement (4.1) by a model of the x_t process and a distributional assumption. Since $x_t < c$ is assumed in the data (2.8) implies

$$x_t = z_t' \gamma + \eta_t, \quad t = 1, \dots, T \quad (4.3)$$

where $z_t' = [1 \ x_{t-1} \ \dots \ x_{t-p}]' = [1 \ X_{t-1}']$ and observations of the presample values x_{-p+1}, \dots, x_0 are supposed to be available. Note that, unlike in the linear case, the regression function in (4.2) does not only depend on the regression coefficients in (4.3) but also on the error variance σ_η^2 . This means that there are rather complicated cross equation restrictions in the model which have to be taken into account in full ML estimation.

For concreteness, the distributional assumption we make in the ML estimation is that the error terms ε_t and η_t have a joint normal distribution. (It is not difficult to see how the subsequent discussion should be modified if other distributions are employed.) Since we saw in the previous section that ε_t and η_t are uncorrelated the normality assumption entails that they are independent Gaussian processes. Thus, the conditional log-likelihood function of the data with x_{-p+1}, \dots, x_0 given is (apart from a constant)

$$\begin{aligned} l_T(\theta) = & -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - f_{t-1}(\psi, \gamma, \sigma_\eta^2))^2 \\ & - \frac{T}{2} \log \sigma_\eta^2 - \frac{1}{2\sigma_\eta^2} \sum_{t=1}^T (x_t - z_t' \gamma)^2 \end{aligned} \quad (4.4)$$

where $\sigma^2 = E\varepsilon_t^2$ and $\theta = [\psi' \gamma' \sigma_\eta^2 \sigma^2]'$ is the vector of all unknown parameters. When a procedure of computing the infinite sum in (2.3) is available it is possible to compute the value of the function $f_{t-1}(\psi, \gamma, \sigma_\eta^2)$ and hence the value of the log-likelihood function $l_T(\theta)$ for any given parameter values. Thus, numerical methods can be used to find a ML estimator of θ . Since analytic derivatives of the function $f_{t-1}^*(\alpha, \gamma, \delta, c, \sigma_\eta^2)$ are not available it is necessary to consider numerical methods which do not require analytical derivatives (or all of them). That the log-likelihood function has first and second order partial derivatives will be demonstrated below.

Since we have assumed that $\eta_t \sim N(0, \sigma_\eta^2)$ and $x_t < c$ for $t = -p+1, \dots, T$ it follows from (3.1) and (3.7) that $E_t x_{t+j}$, $j = 1, 2$, are twice continuously differentiable with respect to the parameters $\mu, \phi_1, \dots, \phi_p, \sigma_\eta^2, \delta$ and c . (Notice that differentiability with respect to c requires the assumption $x_{t-1} < c$ and that differentiability with respect to μ in the case $j = 2$ can readily be shown even though (3.6) involves the a priori assumption $\mu = 0$.) From (3.8) it can further be seen by induction that the same result holds for $E_t x_{t+j} = e_1' K_j(X_t)$ with $j \geq 3$. Thus, the function $f_{t-1}^*(\alpha, \gamma, \delta, c, \sigma_\eta^2)$ defined in (4.1) is twice continuously differentiable from which the same result can be obtained for $f_{t-1}(\psi, \gamma, \sigma_\eta^2)$ and further for the log-likelihood function $l_T(\theta)$.

Thus, since the likelihood function satisfies conventional smoothness conditions and stationary variables are involved one would expect that usual large sample results hold for the ML estimator of θ . This, however, also requires identifiability of the parameter θ as well as more technical conditions, often formulated in terms of uniform convergence of $T^{-1}l_T(\theta)$ and its Hessian. No attempt is made to discuss technicalities of this kind here. From a practical point of view the identifiability issue is of importance but not easy to deal with precisely because no analytic expression of the infinite sum in (4.1) is available. From equations (2.3) and (2.8) it can immediately be seen that identifiability fails if any of the parameters α, β or δ takes the value zero. In the case of α and δ this has already been ruled out by assumption and the same can also be done in the case of β . The parameters γ and σ_η^2 can be uniquely determined from (4.3) so that they are identified while the identifiability of $\alpha, m, \beta, \sigma^2$ seems clear by equations (4.1) and (4.2) and the fact that the conditional expectations in (4.1) do not depend on α . The indentifiability of the parameter δ can be explained by equation (3.9) and the definition of $k_j(X_t)$ in (3.8b). By equation (3.9) and the fact $k_2(X_t) = \Phi(e_1' A X_t - c) / \sigma_\eta$ there seems to be no reason to suspect that the identifiability of the parameter c would fail either. Thus, on the basis of these informal considerations the identifiability of the whole parameter vector θ seems credible.

The above discussion suggests that it is reasonable to apply standard large sample estimation and hypothesis testing results to the model defined by equations (4.2) and (4.3). In particular, if $\hat{\theta}$ denotes the ML estimator of θ its distribution is approximately $N(\theta, (-\partial^2 l_T(\theta) / \partial \theta \partial \theta')^{-1})$ and Wald tests can be constructed in the usual way after the Hessian of $l_T(\theta)$ at $\theta = \hat{\theta}$ has been computed. (This again has to be done numerically.) Of course, corresponding

likelihood ratio and Lagrange multiplier tests can also be used. Two things are worth noting here, however. First, since the error variance σ_η^2 also appears in the regression function (4.2) its ML estimator is not asymptotically independent of the regression coefficient estimators, which is the case in ordinary nonlinear regression models. Second, the null hypothesis $\delta = 0$ cannot be tested in the usual way. The reason is that the threshold parameter c is not identified under this null hypothesis so that usual large sample results do not hold. Special measures are therefore called for the develop a (likelihood based) test in this case (see Hansen, 1996, and the references therein). Of course, the same is also true for the null hypotheses $\alpha = 0$ and $\beta = 0$ but they are hardly of any practical interest. The reason why the hypothesis $\delta = 0$ could be of interest is that it implies that a conventional linear AR model can be used to model the forcing variable x_t and the more complicated TAR model (2.8) is not needed. When no likelihood based test is available for this purpose one may consider the familiar RESET test and less formal procedures like residual plots.

An interesting feature in the present estimation problem is that conventional large sample results are justified even though a threshold model is considered. This is due to the assumption $x_t < c$ (within the sample) which implies the conventional smoothness of the likelihood function. In proper threshold models the likelihood function is not differentiable with respect to the threshold parameter so that the estimation problem becomes nonstandard (see Tong, 1990).

Since the full ML procedure discussed above is iterative it is desirable to have good initial estimates for the parameters. As far as the parameters γ and σ_η^2 are concerned, an obvious way is to apply least squares to equation (4.3). Initial estimates for the parameters $m (= v/(1-\alpha))$, α and β are also easily obtained by a simple GMM procedure to be discussed in more detail in the next section. If initial estimates of the parameters δ and c are also available an initial estimate of the error variance σ^2 can be obtained from the residuals of equation (4.2) in the usual way. Unfortunately, however, there seems to be no simple estimation procedure which could be used to obtain initial estimates for the parameters δ and c , although some rough ideas will be discussed in the next section. If pure initial guesses are employed any ideas of the possible values of δ and c should be made use of. In the case of the parameter c the condition $c > \max \{x_t, t = -p+1, \dots, T\}$ should particularly be taken into account.

5 GMM estimation

Since the ML estimation discussed in the previous section is rather complicated we shall here take a closer look at simpler alternatives based on the GMM approach. The idea is to estimate the parameters v , α and β in (2.1) by applying GMM or, equivalently, instrumental variables estimation to the equation

$$y_t = v + \alpha y_{t+1} + \beta x_t + u_t \quad (5.1)$$

where $u_t = -\alpha(y_{t+1} - E_t^* y_{t+1})$ is a martingale difference sequence satisfying

$$E_t^* u_t = 0 \quad (5.2)$$

Obviously, equation (5.1) is obtained from (2.1) by replacing the unknown conditional expectation $E_t^* y_{t+1}$ by the observable variable y_{t+1} . Although (5.2) implies that the errors in (5.1) are serially uncorrelated it does not rule out conditional heteroskedasticity. In subsequent developments conditional homoskedasticity would not be a necessary assumption but, since it simplifies matters, it will be assumed. Thus, we make the additional assumption that $E_t^* u_t^2 = E u_t^2 = \sigma_u^2$, say.

Now consider the choice of instruments needed in the GMM estimation of the parameters in (5.1). By (5.2) instruments are only needed for the variable y_{t+1} . Valid instruments include x_t and y_t as well as their lagged values and any functions of these. A general form of the GMM estimator can be defined as follows. First, denote $w_t = [1 \ y_{t+1} \ x_t]'$ and let \tilde{y}_{t+1} be the least squares fit in an auxiliary regression of y_{t+1} on chosen instruments collected in the vector q_t . Although not necessary it may be helpful in subsequent discussions to assume that these instruments include 1, x_t and some additional variable(s). Now, if $\lambda = [\nu \ \alpha \ \beta]'$ is the parameter vector to be estimated from (5.1) and $\tilde{w}_t = [1 \ \tilde{y}_{t+1} \ x_t]'$ GMM yields the estimator

$$\tilde{\lambda} = \left(\sum_{t=1}^T \tilde{w}_t w_t' \right)^{-1} \sum_{t=1}^T \tilde{w}_t y_t \quad (5.3)$$

(cf. Hamilton, 1994, p. 420). As is well known and easy to check, the estimator $\tilde{\lambda}$ can be obtained by minimizing the function

$$Q_T(\lambda) = \sum_{t=1}^T (y_t - w_t' \lambda) q_t' \left(\sum_{t=1}^T q_t q_t' \right)^{-1} \sum_{t=1}^T q_t (y_t - w_t' \lambda)$$

Under suitable regularity conditions (including conditional homoskedasticity) the estimator $\tilde{\lambda}$ can be treated as approximately normally distributed with mean value λ and estimated covariance matrix

$$\text{Cov}(\tilde{\lambda}) = \tilde{\sigma}_u^2 \left(\sum_{t=1}^T \tilde{w}_t \tilde{w}_t' \right)^{-1} \quad (5.4)$$

where $\tilde{\sigma}_u^2 = T^{-1} \sum_{t=1}^T (y_t - \tilde{\lambda}' w_t)^2$ is an estimator of σ_u^2 . This result can be used to construct Wald type tests on the parameter vector γ in the usual way.

Above the estimator $\tilde{\lambda}$ implicitly assumed that the instruments used for y_{t+1} were taken from the set $\{1, x_t, x_{t-1}, \dots, x_{t-h}, y_t, y_{t-1}, \dots, y_{t-h}\}$. Since the errors in (5.1) are serially uncorrelated it is possible to use a result of Tauchen (1986) and consider an optimal choice to instruments. Here optimality means obtaining a GMM estimator whose limiting distribution has a minimum covariance matrix

when the relevant information set is that of the econometrician, ie, $\{y_s, x_s; s \leq t\}$. Since conditional homoskedasticity has been assumed Tauchen's (1986) result implies that an optimal choice of instruments is given by the vector $[1 E_t y_{t+1} x_t]'$. From a practical point of view this result is of course infeasible because the conditional expectation $E_t y_{t+1}$ depends on unknown parameters. However, it can still be helpful when one tries to understand the estimation problem and also to find good instruments.

The above discussion implies that an optimal instrument of y_{t+1} is

$$E_t y_{t+1} = m + \beta \sum_{j=0}^{\infty} \alpha^j E_t x_{t+1+j} \quad (5.5)$$

where the equality is based on (2.2). When the forcing variable x_t is assumed to follow the TAR process (2.8) the right hand side of (5.5) also depends on the nonlinear parameters δ and c which, under our assumptions, cannot be estimated from (2.8) alone. Trying to make the optimal instrument of y_{t+1} feasible by using an estimated counterpart is therefore not easy. However, if the values of the parameters δ and c are fixed the optimal instrument can be made feasible by replacing other parameters on the right hand side of (5.5) by estimates discussed in the previous section and simulating the conditional expectations for $j = 1, \dots, J$ and J large. This could be repeated for various values of δ and c and those giving a minimum of the GMM objective function $Q_T(\lambda)$ could be chosen. In this way a rough feasible version of the optimal instrument as well as rough initial estimates of the parameters δ and c can be obtained.

In order to study the above optimal instrument given in (5.5) more closely, suppose again that $\mu = 0$ a priori and note that from (2.3) and (3.9) one readily finds that

$$\begin{aligned} E_t y_{t+1} = m + \frac{\beta}{\alpha} e_1' (I_p - \alpha A)^{-1} X_t - \frac{\beta}{\alpha} x_t \\ + \frac{\beta}{\alpha} \delta e_1' (I_p - \alpha A)^{-1} e_1 \sum_{j=2}^{\infty} \alpha^j k_j(X_t) \end{aligned} \quad (5.6)$$

where the notation is as before. When the instruments used for y_{t+1} include 1, x_t and a third variable the GMM estimator in (5.3) is invariant to nonsingular linear transformations of the vector \tilde{w}_t . From this fact and (5.6) it follows that in the linear case (ie $\delta = 0$) it is optimal to choose the instruments as $\{1, x_t, e_1' (I_p - \alpha A)^{-1} X_t\}$. (Note that this requires the condition $p \geq 2$.) This choice can easily be made feasible because the estimation of the parameters δ and c is not required. Of course, these instruments are also applicable in the nonlinear case (ie $\delta > 0$) although they are then no more optimal because the last term on the right hand side of (5.6) is erroneously ignored. This can be thought of as the effect of the peso problem on the GMM estimation in our set-up. Since no analytic expression of the infinite sum on the right hand side of (5.6) is available it is difficult to give analytic results about this effect on estimation efficiency. Apparently, the size of the parameters δ is a major factor in this respect (to a lesser extent maybe also the size of the parameter c). On the other hand, it is

worth noting that, when $\delta > 0$, instruments of y_{t+1} better than given by $\{1, x_t, e'_1(I_p - \alpha A)^{-1}X_t\}$ may also be found among “conventional” instruments. In particular, since the variables y_t and y_{t-1} also contain information about the nonlinear features of the forcing variable they may provide such instruments.

Since the optimal instrument given by (5.6) is generally rather difficult to apply we shall briefly discuss some simple alternatives which may be of interest. One possibility is to replace the infinite series on the right hand side of (5.6) by an approximation, like a truncated power series or trigonometric series, which is linear in parameters. Terms of such an approximation could then be included in the instrument set to supplement conventional instruments (a feasible version of $e'_1(I_p - \alpha A)^{-1}X_t$ may also be considered here). As the definition of $k_j(X_t)$ in (3.8b) shows, this quantity depends on X_t only through e'_1AX_t and x_t, \dots, x_{t-p+2} so that the above mentioned series approximation could also be based on these variables or maybe only on the first one because, when η_t is assumed to be Gaussian, the “leading” term $k_2(X_t) = \Phi((e'_1AX_t - c)/\sigma_\eta)$ depends on X_t only through e'_1AX_t .

The idea of using a series approximation to approximate optimal instruments is well known and discussed by Newey (1993) in the iid case. Although it may always be possible to find an accurate approximation the number of needed instruments may increase so large that the finite sample properties of the resulting estimator suffer (cf. the simulation results of Tauchen (1986)). Instead of approximating the whole series on the right hand side of (5.6) one might therefore consider using its first term only. Specifically, assuming normality of η_t we have

$$\sum_{j=2}^{\infty} k_j(X_t) = \Phi\left(\frac{e'_1AX_t - c}{\sigma_\eta}\right) + \sum_{j=3}^{\infty} k_j(X_t). \quad (5.7)$$

Replacing A and σ_η by estimates based on (4.3) the first term of the right hand side of (5.7) becomes nonlinear in the parameter c only. Thus, if the latter term on the right hand side of (5.7) is ignored and the value of c is fixed an instrument for y_{t+1} can be based on constant, x_t , $e'_1(I_p - \alpha A)^{-1}X_t$ and $\Phi((e'_1AX_t - c)/\sigma_\eta)$ with α , A and σ_η replaced by previously discussed estimates. Trying various values of c one can then choose the one which minimizes the GMM objective function $Q_T(\lambda)$. Since a search over the values of the single parameter c is only needed here the number of potential values may be larger than in the similar previously discussed procedure where also the parameter δ was involved. Of course, obvious modifications of this approach can be obtained by approximating the series on the right hand side of (5.7) in the same way as discussed in the case of (5.6) or by using also conventional instruments. To facilitate computations, a logistic function or some other approximation may be used to approximate the cumulative distribution function $\Phi(\cdot)$.

One would expect that the approach based on (5.7) is effective when the first term on the right hand side can describe the essential or major part of the nonlinearity in $E_t y_{t+1}$. As a byproduct one also obtains a preliminary estimate for the parameter c . From (5.6) and (5.7) it can be seen that one can also obtain an estimate of $\frac{\alpha}{\beta} \delta e'_1(I_p - \alpha A)^{-1}e_1$. Since α , β and A can here be replaced by

estimates it is also possible to obtain a preliminary estimate for δ . Of course, these preliminary estimates are not efficient and can be very poor if the second term on the right hand side of (5.7) is not negligible. However, they can still be better than purely guessed initial values.

Since the performance of the various GMM estimators discussed in this section cannot be studied analytically the only possibility is to resort to simulation. This is also of interest because the employed optimality arguments are asymptotic and, as seen in Tauchen's (1986) simulation study, optimal instruments may not be the best ones in finite samples. A further motivation for such simulation experiments is to see how much the peso problem described by the TAR process (2.8) actually affects conventional GMM estimators.

We close this section by noting that the discussion given of the GMM estimation may be useful even if the TAR process (2.8) only provides a reasonable approximation of the main characteristics of the peso problem. In particular, the essential point may here be that the potential regime shifts in the forcing variable are related to nonlinearity which is of such a type that the conditional expectations in (2.3) behave roughly in the way implied by the TAR process (2.8). Then optimal instruments of y_{t+1} are also nonlinear functions of the forcing variable and the ideas used in the context of equations (5.6) and (5.7) may prove useful even if the true nonlinearity is not precisely of the type implied by the TAR process (2.8).

6 Simulation results

In order to study properties of the proposed estimators, we perform Monte Carlo simulation experiments. The computing³ difficulties with the maximum likelihood estimation are still unsolved⁴ for the parameter set that we are studying. Hence, we restrict the simulation experiments to study only the GMM estimators.

The processes studied are as follows:

$$y_t = v + \alpha E_t y_{t+1} + \beta x_t \quad t = 1, \dots, T \quad (6.1)$$

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \delta I(x_{t-1} \geq c) + \eta_t \quad \eta_t \sim \text{NID}(0, \sigma^2). \quad (6.2)$$

The parameters of the y_t process take the following values: $v = 0$, $\alpha = 1/1.05$, $\beta = 1$. The process x_t is run with the following parameter values: $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$ (the roots are 0.7 and 0.4), $\delta = \{0, 0.5, 1\}$, $c = 0.193$,⁵ $\sigma^2 = 0.001$. We vary the sample size as follows $T = \{50, 100, 1000\}$. We let x_t run 100 extra

³ The computations were done using Gauss 3.2.37 in 266 MHz Pentium II with Windows NT 4.0 operating system.

⁴ Given the present computer technology and algorithms at hand, the maximum likelihood estimation turned out to be too time consuming for Monte Carlo experiments.

⁵ We use the following formula to determine $c = v/(1-\phi_1-\phi_2) +$

$3 \cdot \sqrt{(1-\phi_2)\sigma^2} / \left[(1+\phi_2) \left((1-\phi_2)^2 - \phi_1^2 \right) \right]$. Based on this value of c , the probability of

regime shift within 100 period is less than 10 per cent.

observations to avoid the impact of the choice of initial values (in our case 0). Note also that we study three choices of δ . The first one, zero, corresponds to the standard, simple linear case; the second one we label as a “mild” peso case and the third one as a “strong” peso case.

In computing the $y_t = v/(1-\alpha) + \beta \sum_{i=0}^{\infty} \alpha^i E_t x_{t+i}$ process we need to forecast the future values of x_t . We follow the procedure suggested by Clements and Smith (1997), which is based on Monte Carlo simulation of future paths of the x_t process. At each point of y_t we replicate the future path of x_{t+i} ($i = 1, \dots, 400$) 500 times.⁶ Note that within the sample the x_t processes involve no regime shifts, ie realizations with $x_t < c$ ($t = 1, \dots, T$) are only accepted. Regime shifts are, however, possible in the forecasts of x_t when the value of the parameter δ in (6.2) differs from zero. Hence future paths of x_t may contain regime shifts. Our forecasting procedure takes this possibility into account. Figure 2 illustrates well how the existence of a peso problem of a TAR form produces time-varying peso premium.

The choice of the instrument set is a crucial part of GMM estimation. We vary the instrument set in our simulations as follows:

1. Constant, x_t , x_{t-1} . Note that in the case of $\delta = 0$ this instrument set is optimal, since the y_t process is a linear combination of the above instruments. We call this *linear instrument set*.
2. Constant, x_t , x_{t-1} , x_t^2 , x_{t-1}^2 , x_t^3 , x_{t-1}^3 . In the second set we augment the first set with second and third power of x_t and x_{t-1} . This is called *polynomial instrument set*.
3. Constant, x_t , $E_t y_{t+1}$. This is an *optimal instrument set* in each case, since it contains the optimal instrument of $E_t x_{t+1}$. As described in the previous sections it is computed from the solution (or numerical approximation of the solution) of Euler equation (6.1). It is based on the true parameter values of the processes and is therefore infeasible in practice.
4. Constant, x_t , $E[y_{t+1} \mid \text{information at } t \text{ and estimated parameter values}]$. The set is as above but the optimal instrument is not based on true parameter values but values estimated from data. The parameters of equation (6.1) are estimated using GMM and the parameter α is restricted to be between zero and unity. A grid search is applied. The linear part of equation (6.2) is estimated by OLS. The value of the parameter c is 0.0001 plus the maximum value of x_t and that of δ is computed as the absolute value of the difference of the mean of y_t and x_t . This is labeled as *estimated optimal instrument set*.
5. Constant, x_t , approximated $E_t y_{t+1}$. In this case the computation of the conditional expectation of y_{t+1} is based on the equation (5.6) with the infinite sum approximated by the first term of the right hand side of (5.7). This means that instead of computing a large number of future values of x_t we compute only two of them. This could give a reasonable approximation to $E_t y_{t+1}$, eg when the value of α is close to zero but not necessarily otherwise. This is called *approximated optimal instrument set*.

⁶ This is the most time consuming part in the computations involved by this study.

In the following simulations we vary four dimensions of the system: We have five choices for instrument sets; $T = \{50, 100, 1000\}$; $\delta = \{0, 0.5, 1\}$; α is estimated freely or restricted between zero and one. Since the parameter δ takes the value zero, our simulations also cover the non-peso case, ie the standard linear case. The simulation results for non-restricted α are summarized in Figures 3–5. In each graph, a smoothed histogram, ie a kernel estimate of the empirical density function of the parameter estimator is computed. This is repeated for different values of the regime shift parameter $\delta = \{0, 0.5, 1\}$. The sample size ($T = \{50, 100, 1000\}$) varies across the rows of the graphs and the instrument set (as described above) across the columns.

The empirical density functions of the estimator of the intercept term, v , are presented in Figure 3. The GMM estimate is fairly precise in the absence of peso problem. The bias is also very small. However, in peso cases the bias may be substantial and uncertainties are large. In these cases the differences between instrument sets are small. It also seems that the use of polynomial instruments (instrument set 2) leads in many cases to marginally smaller biases than the use of other instrument sets.

The most interesting parameter is the discount factor, α . The empirical densities of its estimators are presented in Figure 4. Only in the standard linear case the median is within the feasible range $\alpha \in (0,1)$. Even in such a case there is a high probability of estimating values above one. This demonstrates the failure of achieving a reasonable estimate of the discount factor in many applications of GMM to Euler equation estimation. In the peso case the results are very poor. The median is above one regardless of the choice of the instrument set. This is true even in the largest sample size, $T = 1000$. However, when the sample size is 50 or 100 and the polynomial instrument set is used, the median between zero and one is obtained. The variance of the estimator based on the polynomial instruments is also the smallest among the choices of instrument set. These features are not without cost: They are not preserved in the largest sample size; The bias of the GMM estimator of β is larger with the polynomial instrument set than with other instrument sets. The optimal instrument set (whether based on true or estimated parameter values) has undesirable properties. The variance is large and a major share of the probability mass is located in values above one. This finding is the same as reported by Tauchen (1986). However, the consistency of the optimal instrument is visible, particularly in the standard linear case: the larger the sample size the smaller the variance and the smaller the bias of the estimator.

The problem of obtaining estimates of α above unity might be due to the fact that the GMM estimator, as applied here, does not punish the objective function when the value of the discount factor exceeds unity. To overcome this problem the value of the discount factor may be fixed to an appropriate level or restricted, in course of optimization, into a feasible range.⁷ We experimented⁸ restricting the value of the discount factor between zero and one. This did not solve the problem since typically more than 50 per cent of the probability mass was placed on the (upper) boundary value.

The empirical densities of the estimator of "the fundament effect", ie the parameter β , are reported in Figure 5. The poor performance of the GMM

⁷ This is an "infinite punishment" to the objective function.

⁸ We performed a grid search.

estimator is demonstrated also here. The peso problem leads to a very high variance of the GMM estimator compared to the standard linear case. In the standard, non-peso case, the use of the polynomial instrument set leads to an estimator with a low variance. However, the bias increases substantially. Note that contrary to the other sets this instrument set contains more instruments than the number of parameters. Consequently, this finding corresponds to that of Tauchen (1986): The larger the number of instruments the smaller the variance of the estimator but the larger the bias. The performance of the optimal instruments is satisfactory in the non-peso case.

The results concerning the peso cases are generally not improved when the value of the parameter α is restricted to lie between zero and one. In the peso cases the GMM estimator of β has still undesirable properties. The median is very much higher than the true value. Even the first percentile is above unity. There is, however, one interesting exception: In the standard linear case, the variance of the estimator of β is very low compared to the situation where the value of α is not restricted. Hence, the estimation of parameter β could benefit of restricting the discount factor, α . However, the bias is not a monotonic function of the sample size.

The poor small sample properties of the GMM estimators are not due to the poor choice of model. To measure the goodness of fit, we study the ratio between the variance of the expectational error $u_t \equiv -\alpha(y_{t+1} - E_t y_{t+1})$ and the variance of y_t .⁹ This variance ratio depends only on the values of the parameters of the system. Our choice of parameter values implies the value 0.25 of this ratio¹⁰ in the linear case.

We may summarize the simulation results as follows: First of all, the small sample properties of the GMM estimator – as applied here – are generally poor. It may even be possible that the estimator does not have finite moments for any finite sample size.¹¹ A second feature is that it is very difficult to estimate the discount factor, α , even in the very simple linear case. This might explain results of many empirical studies where GMM is applied to Euler equation estimation: The value of the discount factor is usually fixed. The simulations illustrate that one cannot only blame a wrong model, but rather the estimation method. The restricted estimation leaks the problem into the estimates of parameter β . The third, slightly promising, feature is that the use of polynomial instruments might marginally improve the generally hopeless results.

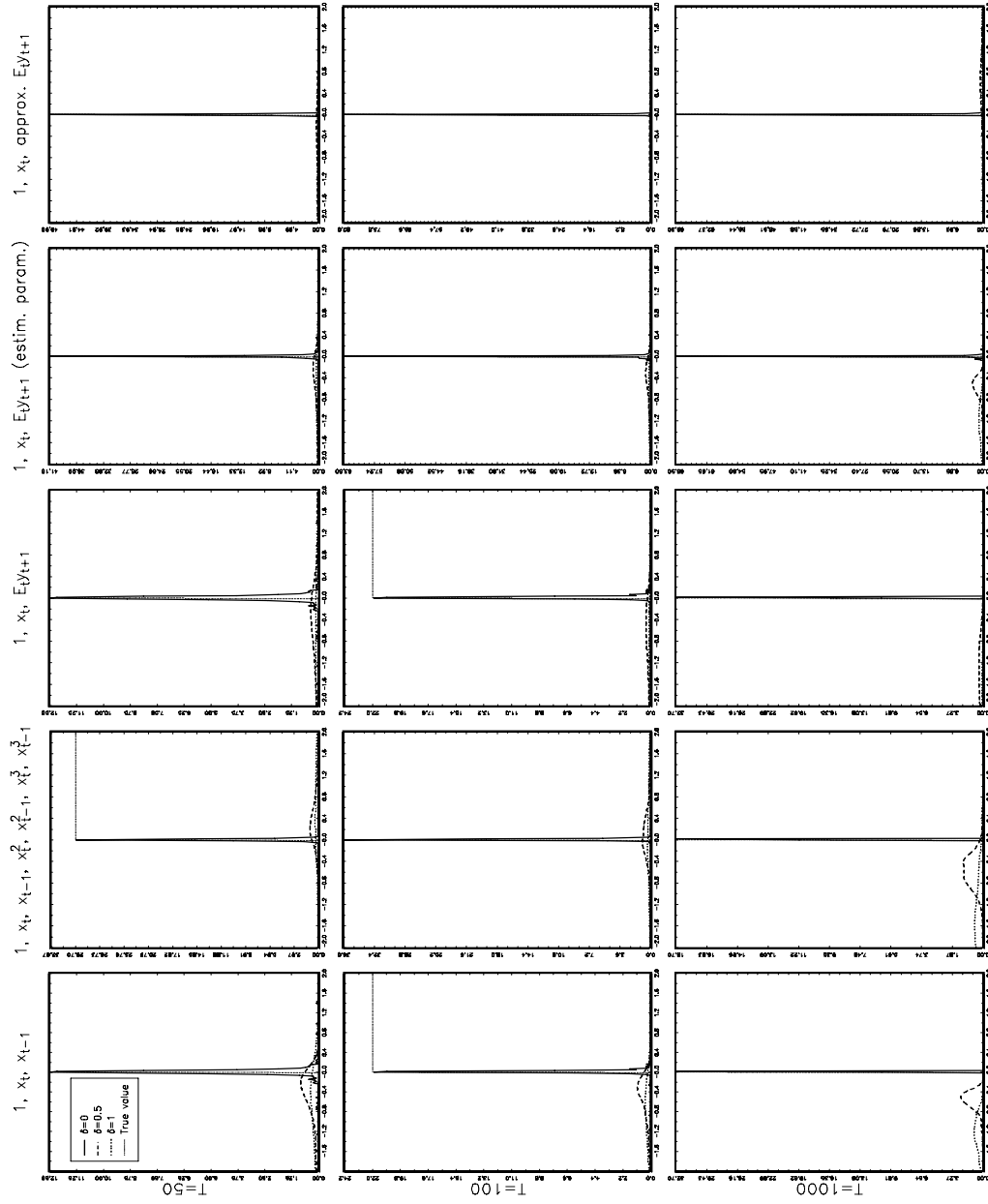
⁹ Note that the multiple correlation coefficient is based on the same variance ratio, $R^2 = 1 - \text{Var}(u_t)/\text{Var}(y_t)$.

¹⁰ This value corresponds $R^2 = 0.75$. Note, however, that in these models R^2 can be negative because y_{t+1} and u_t can be (negatively) correlated.

¹¹ The existence of moments of instrumental variable estimators attracted research up to early 1980s. The main results are summarized, eg, in Davidson–MacKinnon (1993) and Judge et al. (1985). In the case of two-stage least squares, it is known that the number of finite moments (in finite sample sizes) is one less than the number of instruments minus the number of explanatory variables. Hence, in the case when the number of instruments equals the number of explanatory variables, no finite moments exists.

Figure 3.

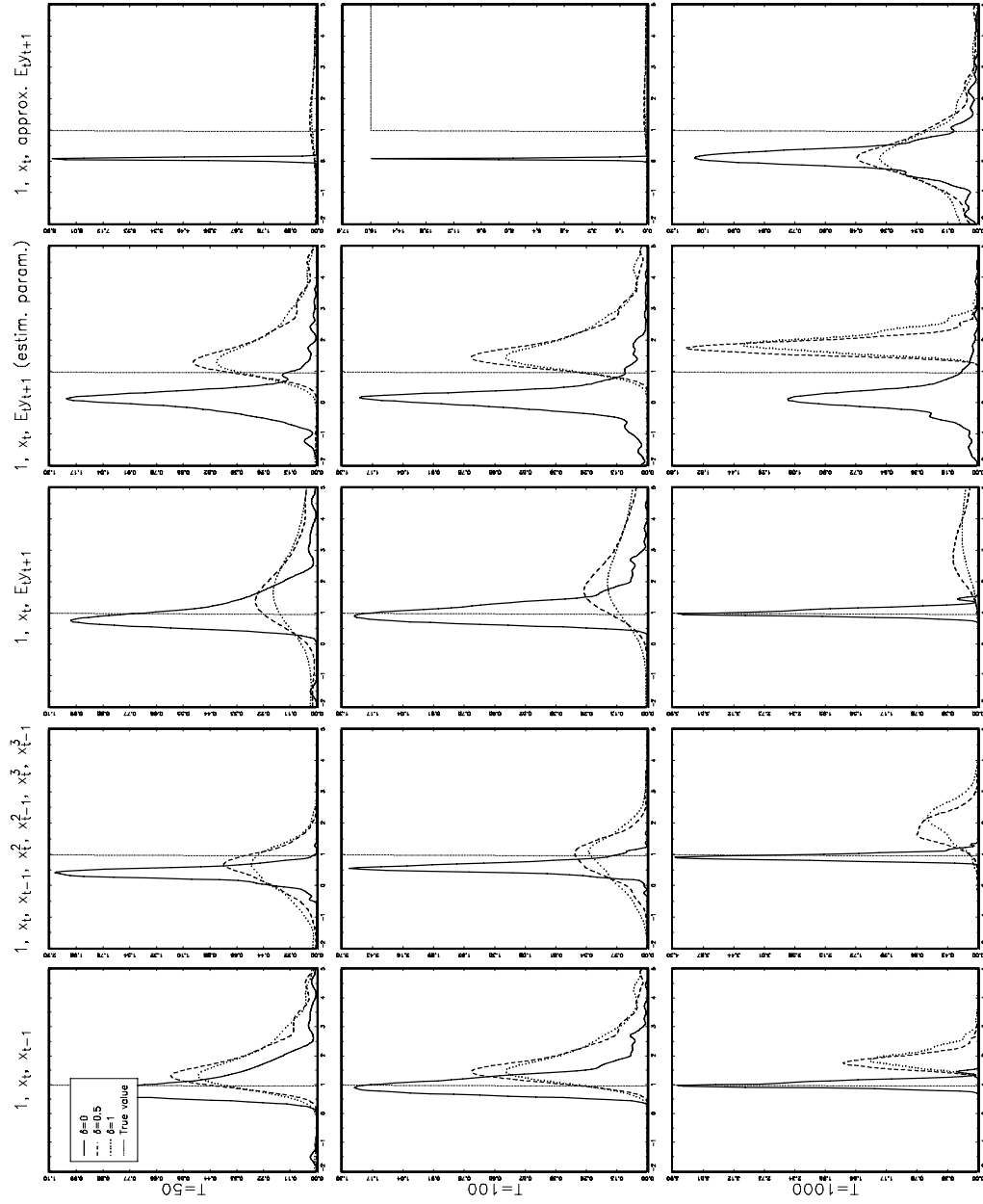
Estimated Density of GMM Estimator of v Based on 500 Draws



The true parameter values are $v = 0$, $\alpha = 1/1.05$, $\beta = 1$, $T = \{50, 100, 1000\}$, $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$ (the roots are 0.7 and 0.4), $\delta = \{0, 0.5, 1\}$, $c = 0.193$ and $\sigma^2 = 0.001$. The instruments sets vary across columns and sample sizes across rows. Monte Carlo simulations are based on 500 draws on normal distribution.

Figure 4.

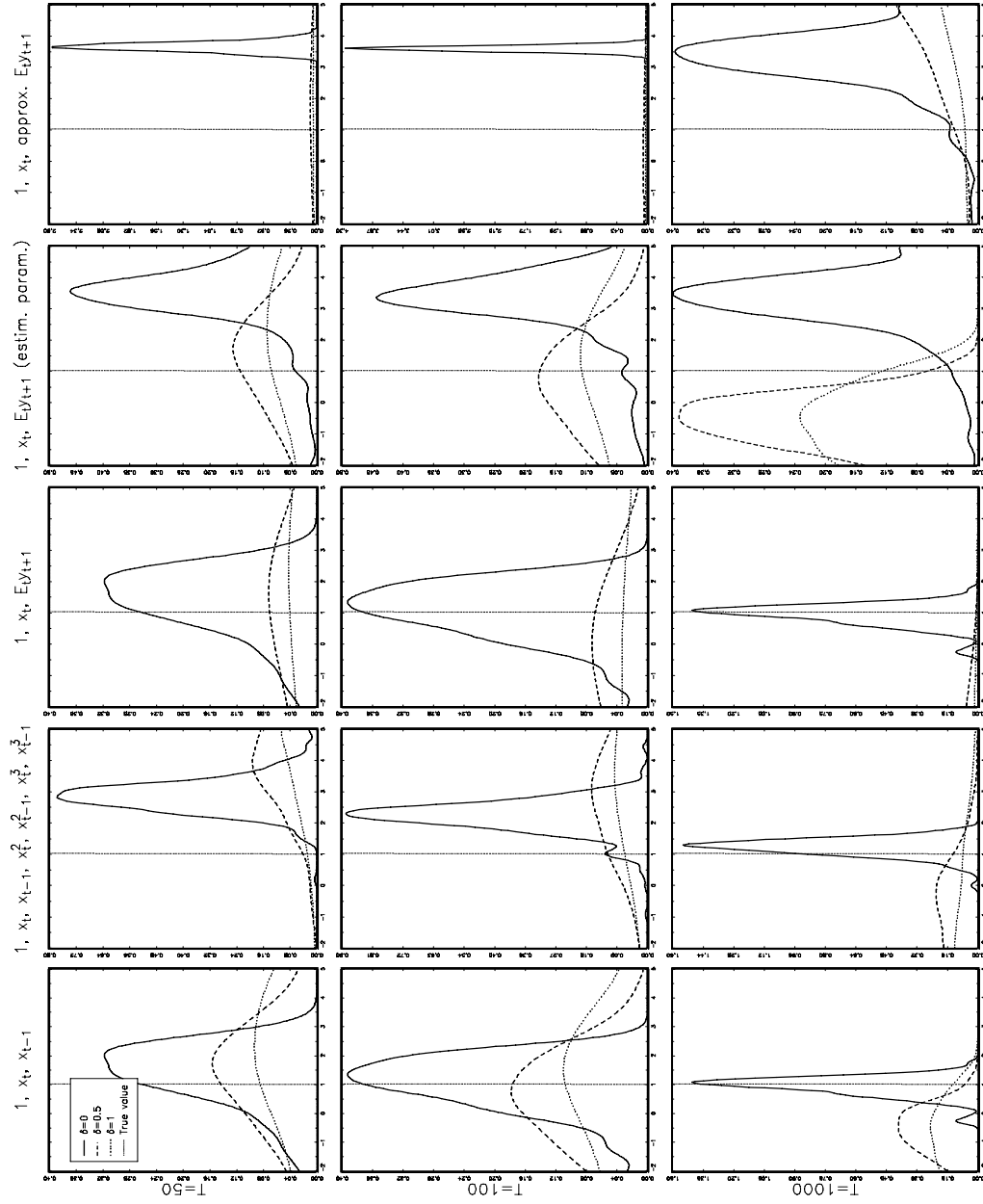
Estimated Density of GMM Estimator of α Based on 500 Draws



The true parameter values are $v = 0$, $\alpha = 1/1.05$, $\beta = 1$, $T = \{50, 100, 1000\}$, $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$ (the roots are 0.7 and 0.4), $\delta = \{0, 0.5, 1\}$, $c = 0.193$ and $\sigma^2 = 0.001$. The instruments sets vary across columns and sample sizes across rows. Monte Carlo simulations are based on 500 draws on normal distribution.

Figure 5.

Estimated Density of GMM Estimator of β Based on 500 Draws



The true parameter values are $v = 0$, $\alpha = 1/1.05$, $\beta = 1$, $T = \{50, 100, 1000\}$, $\mu = 0$, $\phi_1 = 1.1$, $\phi_2 = -0.28$ (the roots are 0.7 and 0.4), $\delta = \{0, 0.5, 1\}$, $c = 0.193$ and $\sigma^2 = 0.001$. The instruments sets vary across columns and sample sizes across rows. Monte Carlo simulations are based on 500 draws on normal distribution.

Table 1.

**Descriptive Statistics when α is not Restricted,
Parameter v**

Sample size	Instru- ment set	δ	True value	Median	Mean	Std. dev.	Bias ^a	RMSE ^b	Bias ^c	MAD ^d
50	1	0	0.000	0.0060	0.0055	0.1598	0.0055	0.1598	0.0060	0.0218
50	2	0	0.000	0.0030	0.0035	0.0161	0.0035	0.0161	0.0030	0.0091
50	3	0	0.000	0.0060	0.0055	0.1598	0.0055	0.1598	0.0060	0.0218
50	4	0	0.000	0.0020	0.0044	0.0558	0.0044	0.0558	0.0020	0.0068
50	5	0	0.000	0.0018	0.0020	0.0084	0.0020	0.0084	0.0018	0.0056
50	1	0.5	0.000	-0.4002	-0.3780	6.9460	-0.3780	6.9460	-0.4002	0.4663
50	2	0.5	0.000	0.2959	0.3149	0.4408	0.3149	0.4408	0.2959	0.3341
50	3	0.5	0.000	-0.4343	-11.4700	267.6000	-11.4700	267.6000	-0.4343	1.1980
50	4	0.5	0.000	-0.4002	-0.3780	6.9460	-0.3780	6.9460	-0.4002	0.4663
50	5	0.5	0.000	-0.3430	-16.6800	204.9000	-16.6800	204.9000	-0.3430	1.0830
50	1	1	0.000	-0.9077	-1.1210	12.6700	-1.1210	12.6700	-0.9077	1.0840
50	2	1	0.000	0.6887	0.8315	1.1850	0.8315	1.1850	0.6887	0.8038
50	3	1	0.000	-0.8934	-18.7600	339.4000	-18.7600	339.4000	-0.8934	2.9580
50	4	1	0.000	-0.9077	-1.1210	12.6700	-1.1210	12.6700	-0.9077	1.0840
50	5	1	0.000	-0.9476	-1.5420	74.6600	-1.5420	74.6600	-0.9476	3.2670
100	1	0	0.000	0.0088	0.0075	0.0718	0.0075	0.0718	0.0088	0.0134
100	2	0	0.000	0.0051	0.0064	0.0126	0.0064	0.0126	0.0051	0.0076
100	3	0	0.000	0.0088	0.0075	0.0718	0.0075	0.0718	0.0088	0.0134
100	4	0	0.000	0.0021	0.0042	0.0427	0.0042	0.0427	0.0021	0.0046
100	5	0	0.000	0.0021	0.0024	0.0050	0.0024	0.0050	0.0021	0.0034
100	1	0.5	0.000	-0.4537	-0.3008	9.5560	-0.3008	9.5560	-0.4537	0.4567
100	2	0.5	0.000	0.0534	0.0790	0.5075	0.0790	0.5075	0.0534	0.2845
100	3	0.5	0.000	-0.8614	-2.0970	93.9100	-2.0970	93.9100	-0.8614	1.3060
100	4	0.5	0.000	-0.4537	-0.3008	9.5560	-0.3008	9.5560	-0.4537	0.4567
100	5	0.5	0.000	-0.7170	5.8600	84.9500	5.8600	84.9500	-0.7170	1.6140
100	1	1	0.000	-1.0760	-2.5140	33.9000	-2.5140	33.9000	-1.0760	1.1020
100	2	1	0.000	0.1161	0.2309	1.2790	0.2309	1.2790	0.1161	0.7555
100	3	1	0.000	-2.3730	-2.7290	54.1800	-2.7290	54.1800	-2.3730	3.8000
100	4	1	0.000	-1.0760	-2.5140	33.9000	-2.5140	33.9000	-1.0760	1.1020
100	5	1	0.000	-2.3130	-5.2880	90.8800	-5.2880	90.8800	-2.3130	4.3730
1000	1	0	0.000	0.0101	0.0101	0.0043	0.0101	0.0043	0.0101	0.0101
1000	2	0	0.000	0.0093	0.0093	0.0040	0.0093	0.0040	0.0093	0.0093
1000	3	0	0.000	0.0101	0.0101	0.0043	0.0101	0.0043	0.0101	0.0101
1000	4	0	0.000	0.0026	0.0062	0.0794	0.0062	0.0794	0.0026	0.0033
1000	5	0	0.000	0.0026	0.0026	0.0155	0.0026	0.0155	0.0026	0.0032
1000	1	0.5	0.000	-0.5500	-0.5766	0.1971	-0.5766	0.1971	-0.5500	0.5499
1000	2	0.5	0.000	-0.5760	-0.5853	0.3162	-0.5853	0.3162	-0.5760	0.5795
1000	3	0.5	0.000	-1.7240	-3.3910	24.4100	-3.3910	24.4100	-1.7240	1.7240
1000	4	0.5	0.000	-0.5500	-0.5766	0.1971	-0.5766	0.1971	-0.5500	0.5499
1000	5	0.5	0.000	0.6296	0.5821	7.1660	0.5821	7.1660	0.6296	0.7324
1000	1	1	0.000	-1.2710	-1.3530	0.5015	-1.3530	0.5015	-1.2710	1.2710
1000	2	1	0.000	-1.5590	-1.5080	0.8179	-1.5080	0.8179	-1.5590	1.5530
1000	3	1	0.000	-4.9940	-6.5220	9.2720	-6.5220	9.2720	-4.9940	4.9980
1000	4	1	0.000	-1.2710	-1.3530	0.5015	-1.3530	0.5015	-1.2710	1.2710
1000	5	1	0.000	1.3420	2.0380	21.3800	2.0380	21.3800	1.3420	1.6400

^aMean – true value^bRoot mean square error^cMedian – true value^dMedian absolute deviation from true value

Table 2.

**Descriptive Statistics when α is not Restricted,
Parameter α**

Sample size	Instrument set	δ	True value	Median	Mean	Std. dev.	Bias	RMSE	Bias	MAD
50	1	0	0.952	0.9160	1.3590	7.2420	0.4065	7.2420	-0.0363	0.3208
50	2	0	0.952	0.3951	0.3781	0.2171	-0.5743	0.2171	-0.5572	0.5561
50	3	0	0.952	0.9160	1.3590	7.2420	0.4065	7.2420	-0.0363	0.3208
50	4	0	0.952	0.1097	0.1372	2.5980	-0.8151	2.5980	-0.8427	0.8595
50	5	0	0.952	0.0690	0.0688	0.0397	-0.8835	0.0397	-0.8834	0.8833
50	1	0.5	0.952	1.5520	1.4930	10.2000	0.5411	10.2000	0.5997	0.7146
50	2	0.5	0.952	0.5897	0.5345	0.6283	-0.4179	0.6283	-0.3627	0.4570
50	3	0.5	0.952	1.6430	18.1600	396.5000	17.2000	396.5000	0.6908	1.7640
50	4	0.5	0.952	1.5520	1.4930	10.2000	0.5411	10.2000	0.5997	0.7146
50	5	0.5	0.952	1.5100	22.3500	247.6000	21.3900	247.6000	0.5572	1.5860
50	1	1	0.952	1.6290	1.7780	8.7380	0.8260	8.7380	0.6764	0.8145
50	2	1	0.952	0.4982	0.3928	0.8494	-0.5596	0.8494	-0.4542	0.5553
50	3	1	0.952	1.6080	13.9700	232.1000	13.0200	232.1000	0.6551	2.1480
50	4	1	0.952	1.6290	1.7780	8.7380	0.8260	8.7380	0.6764	0.8145
50	5	1	0.952	1.7370	1.9530	50.2200	1.0010	50.2200	0.7843	2.3820
100	1	0	0.952	1.0080	1.2040	2.3280	0.2515	2.3280	0.0556	0.2097
100	2	0	0.952	0.5743	0.5925	0.1992	-0.3599	0.1992	-0.3780	0.3790
100	3	0	0.952	1.0080	1.2040	2.3280	0.2515	2.3280	0.0556	0.2097
100	4	0	0.952	0.1435	-0.0818	3.8760	-1.0340	3.8760	-0.8089	0.8613
100	5	0	0.952	0.0700	0.0701	0.0257	-0.8822	0.0257	-0.8824	0.8822
100	1	0.5	0.952	1.6640	1.4440	13.3800	0.4920	13.3800	0.7114	0.7136
100	2	0.5	0.952	0.8910	0.8912	0.7029	-0.0611	0.7029	-0.0614	0.4080
100	3	0.5	0.952	2.2100	4.3210	134.7000	3.3690	134.7000	1.2570	1.8610
100	4	0.5	0.952	1.6640	1.4440	13.3800	0.4920	13.3800	0.7114	0.7136
100	5	0.5	0.952	2.0540	-6.6090	113.2000	-7.5610	113.2000	1.1020	2.3720
100	1	1	0.952	1.7390	2.7920	25.5700	1.8400	25.5700	0.7868	0.8106
100	2	1	0.952	0.9208	0.8218	0.9179	-0.1306	0.9179	-0.0316	0.5274
100	3	1	0.952	2.6430	2.9150	38.4700	1.9630	38.4700	1.6910	2.8230
100	4	1	0.952	1.7390	2.7920	25.5700	1.8400	25.5700	0.7868	0.8106
100	5	1	0.952	2.6370	4.6000	66.4500	3.6470	66.4500	1.6850	3.1540
1000	1	0	0.952	0.9966	1.0230	0.1275	0.0703	0.1275	0.0442	0.0747
1000	2	0	0.952	0.9070	0.9181	0.1037	-0.0343	0.1037	-0.0454	0.0795
1000	3	0	0.952	0.9966	1.0230	0.1275	0.0703	0.1275	0.0442	0.0747
1000	4	0	0.952	0.1076	0.6631	10.6400	-0.2893	10.6400	-0.8448	0.8779
1000	5	0	0.952	0.1151	0.1463	2.1100	-0.8061	2.1100	-0.8373	0.8795
1000	1	0.5	0.952	1.7790	1.8250	0.2743	0.8729	0.2743	0.8266	0.8286
1000	2	0.5	0.952	1.8220	1.8380	0.4477	0.8858	0.4477	0.8696	0.8716
1000	3	0.5	0.952	3.4880	5.7590	33.5300	4.8060	33.5300	2.5350	2.5360
1000	4	0.5	0.952	1.7790	1.8250	0.2743	0.8729	0.2743	0.8266	0.8286
1000	5	0.5	0.952	0.1034	0.1693	10.0100	-0.7831	10.0100	-0.8490	1.0100
1000	1	1	0.952	1.9050	1.9650	0.3550	1.0130	0.3550	0.9522	0.9524
1000	2	1	0.952	2.1100	2.0740	0.5800	1.1220	0.5800	1.1580	1.1580
1000	3	1	0.952	4.5580	5.6230	6.5400	4.6700	6.5400	3.6060	3.6200
1000	4	1	0.952	1.9050	1.9650	0.3550	1.0130	0.3550	0.9522	0.9524
1000	5	1	0.952	0.0489	-0.4388	15.0500	-1.3910	15.0500	-0.9035	1.1310

Table 3.

**Descriptive Statistics when α is not Restricted,
Parameter β**

Sample size	Instru- ment set	δ	True value	Median	Mean	Std. dev.	Bias	RMSE	Bias	MAD
50	1	0	1.000	1.3560	0.3724	20.2600	-0.6276	20.2600	0.3561	1.0740
50	2	0	1.000	2.8650	2.8840	0.5779	1.8840	0.5779	1.8650	1.8650
50	3	0	1.000	1.3560	0.3724	20.2600	-0.6276	20.2600	0.3561	1.0740
50	4	0	1.000	3.5410	3.5060	6.6810	2.5060	6.6810	2.5410	2.6160
50	5	0	1.000	3.6500	3.6560	0.1576	2.6560	0.1576	2.6500	2.6510
50	1	0.5	1.000	1.0120	0.6637	27.6200	-0.3363	27.6200	0.0121	1.8530
50	2	0.5	1.000	4.4080	4.6940	2.5670	3.6940	2.5670	3.4080	3.4130
50	3	0.5	1.000	0.9436	-142.4000	3340.0000	-143.4000	3340.0000	-0.0564	4.0990
50	4	0.5	1.000	1.0120	0.6637	27.6200	-0.3363	27.6200	0.0121	1.8530
50	5	0.5	1.000	1.3340	-53.0200	552.0000	-54.0200	552.0000	0.3339	5.0260
50	1	1	1.000	1.5650	0.7685	46.4700	-0.2315	46.4700	0.5655	3.2230
50	2	1	1.000	6.2890	7.6140	5.5500	6.6140	5.5500	5.2890	5.3270
50	3	1	1.000	1.9000	-85.9300	1630.0000	-86.9300	1630.0000	0.8997	7.6130
50	4	1	1.000	1.5650	0.7685	46.4700	-0.2315	46.4700	0.5655	3.2230
50	5	1	1.000	1.2060	5.9110	219.3000	4.9110	219.3000	0.2062	9.1490
100	1	0	1.000	1.0290	0.4972	6.3970	-0.5028	6.3970	0.0291	0.7271
100	2	0	1.000	2.2300	2.1950	0.5721	1.1950	0.5721	1.2300	1.2310
100	3	0	1.000	1.0290	0.4972	6.3970	-0.5028	6.3970	0.0291	0.7271
100	4	0	1.000	3.4260	3.9670	9.8360	2.9670	9.8360	2.4260	2.4770
100	5	0	1.000	3.6000	3.5990	0.0939	2.5990	0.0939	2.6000	2.6000
100	1	0.5	1.000	0.1022	1.5540	61.9600	0.5535	61.9600	-0.8978	1.5270
100	2	0.5	1.000	3.0450	3.1590	3.0980	2.1590	3.0980	2.0450	2.3020
100	3	0.5	1.000	-1.6570	-2.1710	543.3000	-3.1710	543.3000	-2.6570	5.2070
100	4	0.5	1.000	0.1022	1.5540	61.9600	0.5535	61.9600	-0.8978	1.5270
100	5	0.5	1.000	-1.3210	33.7700	506.2000	32.7700	506.2000	-2.3210	7.3470
100	1	1	1.000	0.1039	-4.7910	113.5000	-5.7910	113.5000	-0.8961	2.4840
100	2	1	1.000	4.4010	4.8900	5.1730	3.8900	5.1730	3.4010	3.9060
100	3	1	1.000	-3.9380	-3.5380	230.1000	-4.5380	230.1000	-4.9380	10.7700
100	4	1	1.000	0.1039	-4.7910	113.5000	-5.7910	113.5000	-0.8961	2.4840
100	5	1	1.000	-3.8270	-21.1500	334.2000	-22.1500	334.2000	-4.8270	12.6100
1000	1	0	1.000	0.9967	0.9307	0.3562	-0.0693	0.3562	-0.0033	0.1946
1000	2	0	1.000	1.2530	1.2200	0.2897	0.2204	0.2897	0.2532	0.2825
1000	3	0	1.000	0.9967	0.9307	0.3562	-0.0693	0.3562	-0.0033	0.1946
1000	4	0	1.000	3.4400	1.9420	29.1500	0.9416	29.1500	2.4400	2.5450
1000	5	0	1.000	3.4350	3.3550	5.8080	2.3550	5.8080	2.4350	2.5540
1000	1	0.5	1.000	-0.6619	-0.8179	1.1670	-1.8180	1.1670	-1.6620	1.6600
1000	2	0.5	1.000	-0.7828	-0.8681	1.8700	-1.8680	1.8700	-1.7830	1.8350
1000	3	0.5	1.000	-7.3260	-17.2200	145.3000	-18.2200	145.3000	-8.3260	8.3190
1000	4	0.5	1.000	-0.6619	-0.8179	1.1670	-1.8180	1.1670	-1.6620	1.6600
1000	5	0.5	1.000	5.9470	5.5510	43.0200	4.5510	43.0200	4.9470	5.3150
1000	1	1	1.000	-1.1230	-1.4500	1.9490	-2.4500	1.9490	-2.1230	2.1230
1000	2	1	1.000	-2.0830	-2.0690	3.2060	-3.0690	3.2060	-3.0830	3.2300
1000	3	1	1.000	-15.3000	-21.3600	35.5000	-22.3600	35.5000	-16.3000	16.3300
1000	4	1	1.000	-1.1230	-1.4500	1.9490	-2.4500	1.9490	-2.1230	2.1230
1000	5	1	1.000	8.8110	11.6500	80.8800	10.6500	80.8800	7.8110	8.4360

7 Conclusions

This paper formalizes the concept of *peso problem* in terms of a linear Euler equation and a nonlinear marginal model. It turns out that the threshold autoregressive model as a marginal model is able to produce time-varying premia – contrary to the widely applied markov switching model. However, due to a nonlinear marginal model, there is no closed form solution to the system of equations. We discuss possible choices of computing the discounted sum of conditional future expectations of the marginal model.

Two estimation strategies emerge. The maximum likelihood is one possibility, but it leads to the numerical approximation of the conditional expectations of the marginal model. This is computationally burdensome and we can not give any Monte Carlo results of the properties of the ML estimator. The second choice is to apply GMM directly to Euler equation estimation. A clear advantage is that no assumption of the marginal model needs to be done. However, the simulation experiments illustrate that the GMM estimator, as applied here, has poor small sample properties in general and particularly so in the peso case.

The main problem with GMM lies in the estimation of the discount factor. The standard way to apply GMM to Euler equation estimation does not punish the GMM criteria when the discount factor gets unfeasible values, eg, values above unity. According to our simulations restricting the discount factor to lie between zero and unity does not solve the problem. The boundary value is attained in more than half of the cases. Improvements in the parameter β are marginal and concern only the non-peso case.

Appendix

Justification of equations (3.8a) and (3.8b)

First note that, for $j = 2$, (3.8a) and (3.8b) follow directly from (3.5) and (3.6). Now suppose that these equations hold for $j \geq 3$ and observe that

$$\begin{aligned} K_{j+1}(X_t) &= \int_{-\infty}^{\infty} K_j(z) p_{X_{t+1}|X_t}(z) dz \\ &= A \int_{-\infty}^{\infty} K_{j-1}(z) p_{X_{t+1}|X_t}(z) dz + \delta e_1 \int_{-\infty}^{\infty} k_j(z) p_{X_{t+1}|X_t}(z) dx. \end{aligned}$$

Here the first equality is based on (3.4). The second one follows from the induction assumption and the fact that, similarly to (3.5), also (3.6) and (3.8) hold even if X_t is replaced by z . By (3.6) the first term is the last expression equals $AK_j(X_t)$. As for the second one, notice that, conditional on X_t , the first component of X_{t+1} has the density function $g(z_1 - e_1' h(X_t))$ while the remaining components have a degenerate distribution at $[x_t \dots x_{t-p+2}]'$. Thus, the last integral equals

$$\int_{-\infty}^{\infty} k_j(z_1, x_t, \dots, x_{t-p+2}) g(z_1 - e_1' h(X_t)) dx_1 = k_{j+1}(X_t)$$

as required.

References

- Abel, A.B. (1982) **Dynamic Effects of Permanent and Temporary Tax Policies in a q Model of Investments.** *Journal of Monetary Economics* 9, 353–373.
- Baker, C.T.H. (1977) **The Numerical Treatment of Integral Equations.** Clarendon Press, Oxford
- Cagan, P. (1956) **The Monetary Dynamics of Hyperinflation.** In M. Friedman, ed., *Studies in the quantity theory of money.* Chicago, University of Chicago Press.
- Clements, M.P. – Smith, J. (1997) **The Performance of Alternative Forecasting Methods for SETAR Models.** *International Journal of Forecasting* 13, 463–475.
- Davidson, R. – MacKinnon, J.C. (1993) **Estimation and Inference in Econometrics.** Oxford University Press, Oxford.
- Hall, R.E. (1978) **Stochastic Implications on the Life Cycle-Permanent Income Hypothesis Theory and Evidence.** *Journal of Political Economy* 86, 971–588.
- Hamilton, J.D. (1993) **Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime.** In G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics.* Vol. 11, New York, North-Holland.
- Hamilton, J.D. (1994) **Time Series Analysis.** Princeton University Press, Princeton.
- Hansen, B.E. (1996) **Inference when a Nuisance Parameter is not Identified under the Null Hypothesis.** *Econometrica* 64, 413–430.
- Judge, G.G. –Griffiths, W.E. – Hill, R.C. – Lütkepohl, H. – Lee, T.C. (1985) **The Theory and Practice of Econometrics.** Second Edition. John Wiley and Sons, New York.
- Masry, E. – Tjøstheim (1995) **Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality.** *Econometric Theory* 11, 258–299.
- Newey, W.K. (1993) **Efficient Estimation of Models with Conditional Moment Restrictions.** In G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics.* Vol. 11, New York, North-Holland.
- Tauchen, G. (1986) **Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data.** *Journal of Business & Economic Statistics* 4, 397–425.
- Tong, H. (1990) **Non-Linear Time Series. A Dynamic System Approach.** Clarendon Press, Oxford.