Maximum Likelihood Estimation of a Unit Root Bilinear Model with an Application to Prices

Daniela Hristova*
Department of Economics
City University London
Northampton Square
London EC1V 0HB, UK.
E-mail: d.hristova@city.ac.uk

June 24, 2004

Abstract

We estimate a unit root bilinear process using the Maximum Likelihood method with log-likelihood function constructed by means of the Kalman filter, and evaluate the finite sample properties of this estimator.

One hundred and six world-wide price series are tested for unit root bilinearity applying the test suggested by Charemza *et al.* (2002b). Applying the Maximum Likelihood estimator based on the Kalman filter, the null hypothesis of no bilinearity is rejected for 40 out of 106 series at the 5% level of significance. Most of the significant unit root bilinear coefficient estimates are explosive.

Keywords: unit root bilinear process, non-linear process, Kalman filter, Simulated Annealing, prices; JEL classification: C13, C22, E31.

1 Introduction

A voluminous literature has been published on the topic of modelling macroeconomic and financial time series, and, as is well known, some of these indicators often appear to follow non-linear processes (see, for example, Hinich and Patterson, 1985, Scheinkman and LeBaron, 1989, among others). A difficulty arising with regard to this issue is to properly specify the form of non-linearity that exists. A class of processes that can (partially) capture possible non-linearity in economic series of data, is the class of bilinear processes (see Granger and Andersen, 1978, Subba Rao and Gabr, 1984, Terdik, 1999, for discussion). Some examples of their applicability to realistic economic and, in particular, financial situations are discussed by Granger and Andersen (1978, pp. 14-16).

Bilinear models are a simple non-linear extension of linear models and have been successfully applied for analyzing macroeconomic and financial series. For example, Byers and Peel (1995), using inter-war exchange rate data, show empirical evidence of bilinearity in the mean, by applying the bilinear-QARCH model, which allows for non-linearity in both mean and variance. Peel and Davidson (1998) propose a bilinear error correction mechanism (ECM) and suggest its application to models in which the "variables display abrupt changes". The authors show that, using annual data on real consumer's expenditure and real gross national product for the UK, the bilinear ECM is superior to both the linear and an alternative non-linear model. On the other hand, Maravall (1983) studies the application of bilinear models for forecasting non-linear processes and demonstrates improvement over ARIMA forecasts when the bilinear model is used.

The present work focuses on a narrow class of bilinear processes, with a single bilinear term, and which, in addition, are 2nd order non-stationary. These processes, called unit root bilinear, have been recently introduced by Charemza, Lifshits and Makarova (forthcoming) (abbreviated CLM hereafter), who have also proposed a two-step testing procedure for unit root bilinearity. Strong evidence of the presence of bilinearity has been established in a large number of mature and emerging stock market indices (see CLM).

Here, we apply the Maximum Likelihood method for estimating the coefficient of bilinearity in unit root bilinear processes. We construct the log-likelihood function using the well-known Kalman filter (KF) algorithm (see Hamilton, 1994, Harvey, 1989, for derivations), which is applicable to non-stationary and non-linear processes alike. The log-likelihood function, formed in this way, is numerically maximized using the Simulated Annealing (SA) algorithm (see Corana et al., 1987, Goffe et al., 1994) and the familiar Newton-Raphson (NR) algorithm. The finite sample properties of the Kalman filter estimator are examined by a series of Monte Carlo simulation experiments and, as an illustration, this technique is applied to a large selection of CPI or RPI world-wide data.

The rest of the paper is organized as follows: Section 2 states the definition and some of the main properties of the unit root bilinear process. Section 3 proposes that the unit root bilinear coefficient be estimated by using the Kalman filter to construct the likelihood, and then maximizing this using either the Simulated Annealing or the Newton-Raphson numerical algorithms. Section 4 presents numerical evidence on the finite sample properties of the maximum likelihood estimator. The two-step testing procedure for unit root bilinearity is outlined in Section 5. A set of 106 world-wide, seasonally unadjusted, monthly price series are examined for bilinearity, and the test and estimation results discussed in Section 6. Finally, Section 7 concludes.

2   Unit root bilinear processes

Let $\{y_t\}_{t=1}^T$ be a discrete time series. A process of the form:

$$y_t = \sum_{p=1}^{P} a_p y_{t-p} + \sum_{q=0}^{Q} c_q e_{t-q} + \sum_{m=1}^{M}\sum_{k=1}^{K} b_{m,k} y_{t-m} e_{t-k}, \tag{1}$$

with $e_t \sim iidN(0, \sigma_e^2)$ and coefficients $c_0 = 1$, $a_p, c_q, b_{m,k} \in \mathbb{R}$, $p = 1, 2, ..., P$, $q = 1, 2, ..., Q$, $m = 1, 2, ..., M$, $k = 1, 2, ..., K$ is called a bilinear process. The bilinear process is denoted by $BL(P, Q, M, K)$ and reduces to linear $ARMA(P, Q)$ iff the coefficients $b_{m,k} = 0$ $\forall\, m, k$.

Detailed theoretical analysis of various bilinear models is presented by Granger and Andersen (1978). We restrict our analysis to investigating the subclass of bilinear processes with a single bilinear term, $BL(1, 0, 1, 1)$. Granger and Andersen (1978) show that the necessary stability condition for the process,

$$y_t = a_1 y_{t-1} + b_{11} y_{t-1}\varepsilon_{t-1} + \varepsilon_t \tag{2}$$

with $a_1, b_{11} \neq 0$, $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$, is $a_1^2 + b_{11}^2 \sigma_\varepsilon^2 < 1$. Introducing the restriction $a_1 = 1$, CLM define the so-called unit root bilinear process, which is our main interest in the present study. We consider a process of the form:

$$\begin{aligned} y_t &= y_{t-1} + b y_{t-1}\varepsilon_{t-1} + \varepsilon_t \\ b &= b_{11} \neq 0, \ \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2). \end{aligned} \tag{3}$$

Clearly, process (3) is non-stationary since it does not satisfy the aforesaid stability condition. Further, the process has a drift that depends on the non-linearity parameter, since $E(y_t) = b\sigma_\varepsilon^2 t$. Furthermore, the first differenced process $\Delta y_t = b y_{t-1}\varepsilon_{t-1} + \varepsilon_t$, is also non-stationary, unlike the first difference of the linear unit root process. To see this, observe that taking $\varepsilon_0 = y_0 = 0$, we find $Var(\Delta y_t) = h(t)$, a function of the time, t (see CLM for details) which does not converge to a constant as $t \to \infty$ and thus the unconditional distribution of Δy_t is not independent of time.

Similarly to Donsker's classical invariance principle, Charemza, Lifshits and Makarova (2002) introduce scaling in space and time of process (3) with factors $1/\sqrt{T}$ and $1/T$, respectively, and show firstly, asymptotic convergence of the scaled process to the solution of a stochastic differential equation and secondly, that this solution (a) is well-defined for $b \in (0, 1/\sqrt{T})$, (b) coincides with the limit process of the linear model for $b \to 0$, and (c) is divergent for $b > 1/\sqrt{T}$. Hereafter, we restrict our analysis to processes of the form (2) with finite sample size T and $b \in \left(0, 1/\sqrt{T}\right)$, which seem to be of practical use in economics, as such unit root bilinear processes are not explosive, with positive mean (recall that $E(y_t) = b\sigma_\varepsilon^2 t$). Although not explosive within sample ($t \leq T$), any such process with constant b coefficient will eventually become explosive, when $T > b^{-2}$, which may be a desirable property when modelling certain price movements.

To visualize consider the sample paths of a few $BL(1, 0, 1, 1)$ processes generated following (3) and using an identical sequence of innovations ε_t, $\varepsilon_t \sim iid\ \ N(0, 1)$ (Fig. 1 (a) - (c) below).

(a) $a = 1$, $b = 0.25/\sqrt{T}$, $T = 1000$ **(b)** $a = 1$, $b = 0.9/\sqrt{T}$, $T = 1000$
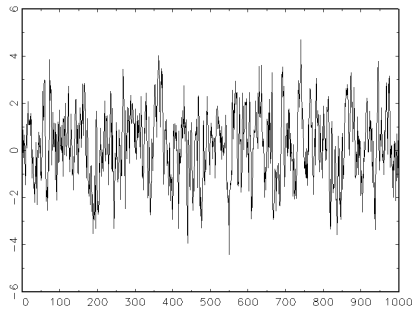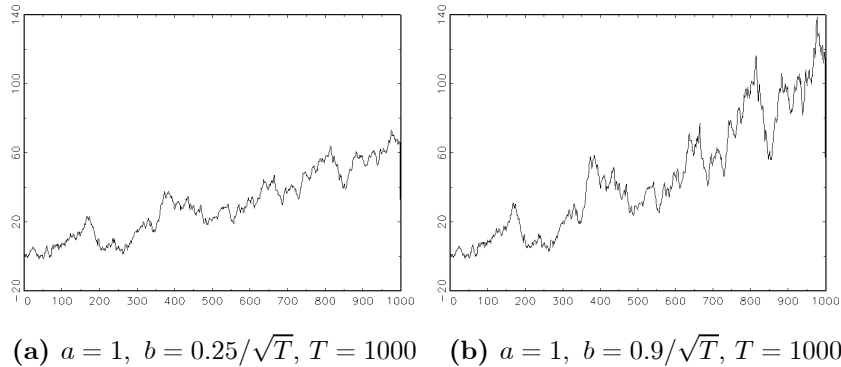
(c) $a = 0.7$, $b = 0.4$, $T = 1000$

Fig. 1 (a)-(b) unit root bilinear processes; **(c)** a stationary bilinear process.

As is evident from Fig. 1 (a)-(b), an increase in the coefficient of bilinearity induces larger variability in the generated processes as well as a steeper trend in the mean.

3 Estimating the bilinear coefficient in a unit root bilinear process

In our univariate, unit root bilinear process, (3), the coefficient of bilinearity b determines the impact of the non-linear part of this process, $y_{t-1}\varepsilon_{t-1}$, with $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$, $t = 1, 2, ..., T$, on the observed variable y_t. The non-linear part contains the innovations ε_{t-1}, which are unobserved, thus leading to difficulties with regard to the estimation of the bilinear coefficient. The present work suggests estimating process (3) by means of the Maximum Likelihood method. We construct the log-likelihood function, $\ln F(y; b, \sigma_\varepsilon^2)$, applying the Kalman filter to the state-space form presented in Section 3.1:

$$\ln F(y; b, \sigma_\varepsilon^2) = \ln \prod_{t=1}^{T} f_t(y_t | \Psi_{t-1}) = \sum_{t=1}^{T} \ln f_t(y_t | \Psi_{t-1}),$$

where b and σ_ε^2 stand for the unknown parameters, $\Psi_t = \{y_0, y_1, y_2, ..., y_t\}$ is the set of observations available at time $t = 1, 2, ..., T$, and $f_t(y_t | \Psi_{t-1})$ denotes the recursively expressed probability density function of y_t given Ψ_{t-1}. Recursive substitution of $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_1$ (i.e. $\varepsilon_{t-m} = \Delta y_{t-m} - b y_{t-(m-1)} \varepsilon_{t-(m-1)}$, $m = 1, ..., t-1$) into equation (3) yields:

$$y_t = y_{t-1} + \sum_{j=1}^{t-1} (-1)^{j+1} b^j \Delta y_{t-j} \cdot \prod_{k=1}^{j} y_{t-k} + b^t \varepsilon_0 \prod_{i=1}^{t} y_{t-i} + \varepsilon_t. \qquad (4)$$

Since $0 < b < 1/\sqrt{T}$ the term $b^t \varepsilon_0 \prod_{i=1}^{t} y_{t-i}$ vanishes with increasing t and, given that $\{\varepsilon_t\}_{t=0}^{T} \sim iidN(0, \sigma_\varepsilon^2)$, the distribution of y_t conditional on Ψ_{t-1} is approximately Gaussian with known mean and variance (see Hamilton, 1994, for the corresponding mean and variance expressions). Note that if we take $\varepsilon_0 y_0 = 0$ as an initial condition then the conditional distribution is exactly Gaussian. In addition, since equation (4) leads to a log-likelihood function, which is a polynomial of degree $2t$ with respect to b, $\ln F(y; b, \sigma_\varepsilon^2)$ may have more than one real local maximum.

3.1   The Kalman Filter algorithm

To construct the log-likelihood function, $\ln F(y; b, \sigma_\varepsilon^2)$, we convert the unit root bilinear process into a state-space form (i.e. a system of two equations: state equation and observation equation), and apply a tool applicable to non-linear and non-stationary processes, namely the Kalman filter algorithm. It is, however, well-known that the state-space form of a given process is not necessarily unique (see, for example, Harvey, 1989, pp. 102-103). Moreover, the non-linearity of our underlying process (3) might result in state-space representations with a state vector following a non-stationary process. This would lead to difficulties with regard to the choice of the state vector initial values, so we use the following state-space representation of (3), which is computationally convenient as the state vector is a stationary process (see below):

$$
\begin{aligned}
y_t &= \mathbf{A}_{t-1} + \mathbf{B}_{t-1} \boldsymbol{\eta}_t & \textit{observation equation} \\
\boldsymbol{\eta}_{t+1} &= \mathbf{F} \, \boldsymbol{\eta}_t \, + \, \boldsymbol{\omega}_{t+1} & \textit{state equation}
\end{aligned}
$$

where $t = 1, 2, ..., T$, $\mathbf{A}_{t-1} = [y_{t-1}]$ and $\mathbf{B}_{t-1} = \begin{bmatrix} b y_{t-1} & 1 \end{bmatrix}$ are time-varying coefficient matrices, $\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and, $\boldsymbol{\eta}_{t+1} = \begin{bmatrix} \varepsilon_t & \varepsilon_{t+1} \end{bmatrix}'$ and $\boldsymbol{\omega}_{t+1} = \begin{bmatrix} 0 & \varepsilon_{t+1} \end{bmatrix}'$ with $\{\varepsilon_t\}_{t=0}^{T} \sim iidN(0, \sigma_\varepsilon^2)$. The coefficient of bilinearity, b, and the variance of the innovations, σ_ε^2, are the parameters to be estimated. Given the set of observations $\Psi_t = \{y_0, y_1, y_2, ..., y_t\}$ available at time t, the Kalman filter algorithm recursively generates an optimal non-linear forecast of the state vector $\boldsymbol{\eta}_{t+1}$, $t = 1, ..., T$. Because the state vector $\boldsymbol{\eta}_t$ is a stationary process, we may start the algorithm with an initial value $\hat{\boldsymbol{\eta}}_{1|0} = \begin{bmatrix} 0 & 0 \end{bmatrix}'$ (see Hamilton, 1994). In brief, the KF algorithm consists in the consecutive application of the following

four steps, executing steps 2 to 4 T times:

> *step 1*: initialization of the state vector $\hat{\boldsymbol{\eta}}_{1|0}$;
>
> *step 2*: optimal least squares forecast of the next observation $\hat{y}_{t|t-1}$, $t = 1, ..., T$, based on the information available at time $t - 1$, namely the conditional distribution of y_t;
>
> *step 3*: updating the state vector $\hat{\boldsymbol{\eta}}_{t|t}$, $t = 1, ..., T$, based on the information available at time t;
>
> *step 4*: one period ahead optimal forecast of the state vector $\hat{\boldsymbol{\eta}}_{t+1|t}$, $t = 1, ..., T$, based on the information available at time t.

We estimate the unknown parameters b and σ_ε^2 applying the Maximum Likelihood method to the string of conditional distributions delivered at Step 2. It is straightforward to show that the log-likelihood function $\ln F(y; b, \sigma_\varepsilon^2)$, derived using the KF algorithm applied to the state-space form above, is bounded for $b < 1/\sqrt{T}$. To maximize $\ln F(y; b, \sigma_\varepsilon^2)$ we require a numerical optimization algorithm; however, some such methods (e.g. BFGS, BHHH, DFP, Polak-Ribiere Conjugate Gradient), often fail to maximize our log-likelihood function.

3.2 Maximizing the log-likelihood function

If speed is required then, by sacrificing some accuracy, we maximize $\ln F(y; b, \sigma_\varepsilon^2)$ by the well-known Newton-Raphson (NR) algorithm; however, when precision is needed or NR fails to maximize our log-likelihood function, we adopt the so-called Simulated Annealing (SA) algorithm (see Corana et al., 1987, Goffe et al., 1994, among others).

The SA algorithm has a number of advantages over the conventional numerical optimization techniques. It has fewer inherent limitations, for example there is no need for the function to be smooth or continuous, and it is applicable to multivariate, ill-conditioned functions with many local optima. It overcomes the main problems associated with traditional optimization algorithms (e.g. infinitely large parameter values, looping through the same point, slow convergence or lack of convergence, finding local instead of global optimum, etc.) and finds the global optimum or a good, near-optimal local optimum. Corana et al. (1987), for example, test the SA algorithm against (a) the Nelder and Mead simplex method (see Nelder and Mead, 1965) and (b) a global optimizer using Adaptive Random Search (see Marsi et al., 1980), applying these methods to the Rosenbrock function (of dimensions 2 and 4) and to multiminima functions (of dimension 2, 4 and 10). In all the cases SA found the global minimum or the local minimum closest to the global one and, with a few exceptions, proved to be "much more reliable and efficient than the other algorithms" (see Corana et al., 1987, pp. 278). Later, Goffe et al. (1994) apply SA to four econometric problems and compare the SA's performance to the one of few conventional algorithms (i.e. simplex, conjugate gradient and quasi-Newton algorithms), showing that the SA algorithm is superior.

A major advantage of the SA algorithm is its ability to find the global optimum irrespective of the initial parameter values. As is well-known the conventional numerical optimization techniques, in particular the NR algorithm, require suitably chosen initial values, which should be as close as possible to the true parameter values. On the other hand, a disadvantage of the SA algorithm is its high computational cost (see Goffe et al., 1994). As a result, in our simulation studies (see Section 4) we use the NR algorithm. When applying NR for maximizing the log-likelihood function, $\ln F(y; b, \sigma_\varepsilon^2)$, we assign b_0 and σ_0^2 -

the starting values of the unknown parameters b and σ_ε^2 - using initial estimates constructed similarly to a suggestion of CLM: under the assumption that the bilinear coefficient is approximately zero, i.e. $b \approx 0$, equation (3) is equivalent to $\Delta y_t \approx \varepsilon_t$; consequently, applying the OLS method to $\Delta y_t = by_{t-1}\Delta y_{t-1} + \varepsilon_t$, allows one to roughly estimate the coefficient of bilinearity, giving the estimate, $\widehat{b}_{OLS}$. Then, with $b_0 = \widehat{b}_{OLS}$, we estimate the variance from the residuals, giving, $\sigma_0^2 = \hat{\sigma}_{\widehat{\varepsilon}_{OLS}}^2$. If the NR algorithm fails to maximize the log-likelihood function, we switch to the SA algorithm. SA is also used in the application to prices (see Section 6).

In brief, whilst the conventional optimization algorithms (including NR) move uphill iteratively, by moving uphill and downhill the SA algorithm ensures first, independence from the initial parameter values and second, the ability to escape from a local optimum and to proceed to find the global one. The Simulated Annealing approach is described in the following section.

3.3 The Simulated Annealing Algorithm

Intuitively, the principles of the SA algorithm can be compared with those of a physical process by which molten metal is cooled. If the process of cooling is slow (known as annealing), the metal passes gradually from high to low energy state, that is, to the global minimum energy state of the system. If, however, the metal is cooled rapidly, when fully cooled it might contain more energy than annealed metal, that is the system will be in a local minimum state with higher energy than the energy of a slowly cooled metal system.

Assume one is interested in finding the global maximum of a bounded function, $L(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$ is an n dimensional vector collecting the parameters to be estimated (e.g. in our case $L(\boldsymbol{\theta}) = \ln F(y; \boldsymbol{\theta})$, with $\boldsymbol{\theta} = \begin{bmatrix} b & \sigma_\varepsilon^2 \end{bmatrix}'$). We outline the SA algorithm following the work of Corana et al. (1987) and Goffe et al. (1994). The algorithm consists in the successive execution of the following steps:

step 1: Initialization of the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$, and the temperature $\overline{T}$:
· $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ is an n dimensional vector of parameters such that the i-th coordinate, θ_i^0, $i = 1, ..., n$, takes any value in an initially specified interval, in which the parameter θ_i lies;
· $\boldsymbol{\nu} = \boldsymbol{\nu}^0$ is an n dimensional vector of step lengths such that the i-th coordinate, ν_i^0, $i = 1, ..., n$, corresponds to θ_i's initial step length;
· temperature $\overline{T} = \overline{T}^0$;
step 2: A new point, $\boldsymbol{\theta}_{new} = (\theta_1, ..., \theta_{i,new}, ..., \theta_n)$, is generated by the rule:

$$\theta_{i,new} = \theta_i + r.\nu_i$$

where $i = 1, \ldots, n$ is the i-th coordinate direction and $r \in [-1, 1]$ is a uniformly distributed random number;
step 3: A decision of acceptance or rejection of the new point, $\boldsymbol{\theta}_{new}$, is formed as follows:
· if $L(\boldsymbol{\theta}_{new}) \geqq L(\boldsymbol{\theta})$ then $\boldsymbol{\theta}_{new}$ is accepted, $\boldsymbol{\theta} := \boldsymbol{\theta}_{new}$, and the algorithm moves uphill;
· if $L(\boldsymbol{\theta}_{new}) < L(\boldsymbol{\theta})$ then $\boldsymbol{\theta}_{new}$ is accepted with probability

$$P = e^{(L(\boldsymbol{\theta}_{new}) - L(\boldsymbol{\theta}))/\overline{T}}. \tag{5}$$

If P is greater than a uniformly distributed random number P', $P' \in [0,1]$, then $\boldsymbol{\theta} := \boldsymbol{\theta}_{new}$ and the algorithm moves downhill. This is known as the Metropolis criterion (see Metropolis et al., 1953);

step 4: Steps 2 to 3 are repeated for each coordinate direction i, $i = 1, ..., n$;

step 5: Steps 2 to 4 are repeated N_ν times;

step 6: The step vector, $\boldsymbol{\nu}$, is adjusted such that one-half of the total number of moves are accepted (see Corana et al., 1987, for more detail);

step 7: Steps 2 to 6 are repeated $N_{\overline{T}}$ times, that is the length of the step vector is updated $N_{\overline{T}}$ times;

step 8: The temperature is reduced following the rule:

$$\overline{T}_{new} = r_{\overline{T}}.\overline{T},$$

where the reduction factor $r_{\overline{T}} \in [0,1]$ is set by the user; $\overline{T} := \overline{T}_{new}$. Clearly, the higher the coefficient $r_{\overline{T}}$ the slower the temperature falls;

step 9: Steps 2 to 8 are repeated until a termination criterion is satisfied (see Corana et al., 1987, for more detail).

It is clear from (5) that starting at the current maximum and at lower temperature decreases the number of downhill moves and, consequently, the length of the step vector, $\boldsymbol{\nu}$, declines, thus, concentrating the new search on the most promising area. The temperature is reduced every $N_\nu.N_{\overline{T}}$ cycles of moves along every direction and after $N_{\overline{T}}$ step adjustments, that is till it is low enough so that no useful improvement can be expected from further temperature diminishing. The parameters N_ν, $N_{\overline{T}}$ and $r_{\overline{T}}$ are set by the user and suggestions for their values are presented in Corana et al. (1987).

When using the SA algorithm, however, one needs to carefully choose the initial temperature $\overline{T}^0$ (see Corana et al., 1987). Detailed discussion of this issue is also presented in the work of Goffe et al. (1994), who propose a useful way of determining the initial temperature, $\overline{T}^0$, namely: starting at low temperature, say, $\widetilde{T}^0 = 10$, and with reduction factor larger than unity, say, $\widetilde{r}_{\widetilde{T}} = 1.5$, we identify the temperature, $\widetilde{T}$, at which all the n intervals in which the unknown parameters must lie, are in the search area. The initial temperature is then set as $\overline{T}^0 = \widetilde{T}$.

4   Finite sample properties of the Kalman filter estimator

To examine the finite sample *bias* and *rmse* of the KF estimator applied to a unit root bilinear process, we conducted a series of Monte Carlo simulation experiments for samples of size $T = 50, 100, 250, 500$, and coefficients of bilinearity b in the range $\left(0, 1/\sqrt{T}\right)$. In each experiment we generate series of length $T + 100$, following a DGP of the form (3) with $\varepsilon_t \sim iid\ N(0,1)$; discard the first 100 elements, apply the NR algorithm to the likelihood obtained via the Kalman filter to obtain the estimates of b and repeat $1,000$ times to calculate *bias* and *rmse* of the KF estimator. The NR algorithm is initialized using the true parameter values.

Table 1 *Bias* and *rmse* of the Kalman filter estimator

for samples of size $T = 50, 100, 250, 500$ and $b \in (0, 1/\sqrt{T})$;

number of Monte Carlo replications: 1,000.

T	50		100		250		500	
b	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*
0.01	-0.0009	0.0367	0.0004	0.0219	0.0001	0.0082	0.0001	0.0040
	(-0.025)		(0.014)		(0.012)		(0.025)	
0.02	-0.0015	0.0357	0.0005	0.0201	0.0002	0.0079	0.0000	0.0040
	(-0.042)		(0.017)		(0.025)		(0.000)	
0.03	-0.0012	0.0342	0.0003	0.0190	0.0003	0.0079	0.0000	0.0045
	(-0.035)		(0.011)		(0.038)		(0.000)	
0.04	-0.0003	0.0328	-0.0001	0.0180	0.0006	0.0092	0.0000	0.0063
	(-0.009)		(-0.004)		(0.065)		(0.000)	
0.05	0.0001	0.0340	-0.0001	0.0167	0.0003	0.0121		
	(0.003)		(-0.003)		(0.025)			
0.07	0.0001	0.0390	0.0001	0.0212				
	(0.003)		(0.003)					
0.09	0.0001	0.0390	-0.0004	0.0303				
	(0.003)		(0.012)					
0.12	0.0025	0.0548						
	(0.046)							

Table 1 above presents the results of this empirical exercise. The *t*-statistics given in brackets reveal insignificance of the bias at the 10% level of significance. It is evident that for fixed $b \in \left(0, 1/\sqrt{T}\right)$, the *rmse* of the KF estimator progressively decreases with an increase in the sample size. With a few exceptions, similar conclusion holds for the *bias* of the KF estimator. For fixed sample size T, however, a clear increase in the *rmse* values is observed for b values approaching the theoretical upper limit of $1/\sqrt{T}$. [1]

To test the significance of the unit root bilinear coefficient b estimated by the KF algorithm, we calculate the Student *t*-statistic critical values for samples of different size,

---

[1]The empirical exercise described above was repeated for the OLS estimator. Comparison between the OLS and the KF estimator reveals that (a) the *bias* of the OLS estimator is negative and larger in absolute value, and (b) the OLS estimator has smaller *rmse* for samples of small size and bilinearity of small magnitude, more precisely for $\{T = 50; b = 0.01, ..., 0.05\}$, $\{T = 100, b = 0.01, ..., 0.03\}$, $\{T = 250, b = 0.01\}$.

$T = 50, 100, 250, 500$, and levels of significance 0.01, 0.05, 0.1 (see Table 2 below). We generate 30,000 random walk series $y_t = y_{t-1} + e_t$ with $e_t \sim iid\ N(0,1)$, $t = 1, 2, ..., T$, each time estimating a regression of the form $\Delta y_t = by_{t-1}e_{t-1} + e_t$. Because the SA algorithm is very slow, this experiment again uses the NR algorithm to maximize the log-likelihood function. The parameters b and σ_ε^2 are initialized as explained in Section 3.2 above. If NR fails, we apply the SA algorithm. In addition, to allow for a drift different from $b\sigma_\varepsilon^2 t$, we also calculate the Student t-statistic critical values for regressions of the form $\Delta y_t = d + by_{t-1}e_{t-1} + e_t$, where d is a constant.

Table 2 Critical values of the Kalman filter Student t-statistic;
number of Monte Carlo replications: 30,000.

| Sample size T | Levels of significance | | | | | |
| | $d = 0$ | | | $d \neq 0$ | | |
	0.01	0.05	0.1	0.01	0.05	0.1
50	3.2393	1.9514	1.4454	3.3662	2.0202	1.4931
100	2.6781	1.7861	1.3710	2.7468	1.8205	1.3881
250	2.4764	1.7108	1.3268	2.4469	1.7093	1.3108
500	2.3535	1.6611	1.2794	2.3666	1.6481	1.2698
Normal distribution (∞)	2.3263	1.6449	1.2816	2.3263	1.6449	1.2816

As is evident from Table 2, the distribution of the Kalman filter estimates shrinks with an increase in sample size, approaching the Normal distribution for samples of large size.

5 Testing for unit root bilinearity

Clearly, the distribution of the bilinear coefficients in process of the form (2) would vary with the variation in the values of the coefficient a_1. Since in the present study we investigate the class of processes (2) with $a_1 = 1$ only (i.e. process (3)), before testing for bilinearity (i.e. $b \neq 0$) we test for presence of a unit root in the first-order (linear) autoregressive process. This is the first step of the two-step testing procedure suggested by CLM. Conditional on the presence of a unit root, the second step consists in testing the null hypothesis of no bilinearity, $b = 0$, versus the alternative of bilinearity, $b \neq 0$, in a regression of the form $\Delta y_t = by_{t-1}\varepsilon_{t-1} + \varepsilon_t$ (for more detail regarding the validity of the conditional, unit root bilinear testing procedure see CLM). Clearly, under the null $\Delta y_t = by_{t-1}\Delta y_{t-1} + \varepsilon_t$ with $b = 0$. Based on the OLS method, CLM derive the Student t-test statistic for estimated models with and without a constant (i.e. $\Delta y_t = (d) + by_{t-1}\Delta y_{t-1} + \varepsilon_t$), and prove convergence to functionals of Wiener processes, which are $N(0,1)$. They report finite sample critical values for both tests and show that the distributions quickly converge to the Normal. Investigation of the power of these tests shows that, for $b \in \left(0, 1/\sqrt{T}\right)$, power rises monotonically with increasing b, whilst for $b \geq 1/\sqrt{T}$ it declines quickly. Moreover, for samples of size 200 or larger and b values close to $1/\sqrt{T}$, the power of the test is close to unity.

In the practical exercise below we test for the presence of bilinearity applying the OLS method (with critical values calculated for the purposes of this analysis) and the Maximum Likelihood method based on the KF (with critical values reported in Table 2), and compare the results. Rejection of the null hypothesis $b = 0$ leads to an inference in favour of bilinearity.

6 Application to prices

6.1 Data and results

The data are monthly, seasonally unadjusted consumer or retail price indices (CPI or RPI) for 106 countries, collected from Datastream (see www.datastream. com). The series vary in length from 120 to 372 observations, and cover various time periods between January 1970 and December 2001.

All the 106 price series are tested for a bilinear unit root at the 5% level of significance. Initially, we test for a unit root applying the well-known DF test (MacKinnon, 1991, critical values). If the null of a unit root is not rejected we proceed with testing for bilinearity (see Section 5). We reject the null hypothesis of no bilinearity for 37 price series, if the OLS estimator is applied, and for 40 price series, if the KF estimator is used. All the estimated regressions include a constant. A summary of the test and estimation results is presented in Table 3, which lists the significant bilinear coefficient estimates only. To find the KF estimates of the bilinear coefficient, $\widehat{b}_{KF}$, we used the SA algorithm[2] with $N_\nu = 20$, $N_{\overline{T}} = 20$ and $r_{\overline{T}} = 0.75$. The rest of the parameters were set to the values suggested by Corana et al. (1987).

---

[2] The Simulated Anealing algorithm was coded in GAUSS by E.G.Tsionas.

Table 3 Testing for unit root bilinearity in prices;
estimated regressions with a constant; level of significance: 0.05.

			DF	OLS		KF	
country name	T	$1/\sqrt{T}$	t-st.	$\widehat{b}_{OLS}$	t-st.	$\widehat{b}_{KF}$	t-st.
Aruba	170	0.077	-1.29	-0.061	-3.74	-0.069	-4.29
Botswana	306	0.057	-2.14	——^{a}	——^{a}	0.035	2.18
Bulgaria	132	0.087	-2.65	0.048	4.22	0.043	3.89
Chad	214	0.068	-0.67	0.056	3.71	0.056	3.75
Colombia	352	0.053	-1.52	0.064*	3.74	0.105*	3.24
Costa Rica	312	0.057	-1.21	0.154*	8.22	0.133*	7.68
Dominican Rep.	378	0.051	0.34	0.165*	11.70	0.128*	10.60
Ecuador	331	0.055	3.10	0.121*	5.79	0.116*	7.47
El Salvador	371	0.052	-0.74	0.092*	5.65	0.074*	4.95
Ethiopia	378	0.051	-1.56	0.031	2.28	0.029	2.18
Gambia	368	0.052	-1.19	0.048	2.78	0.045	2.67
Ghana	356	0.053	-1.48	0.112*	7.50	0.136*	8.86
Greece	360	0.053	-2.31	——^{a}	——^{a}	0.062*	2.17
Guatemala	371	0.052	0.65	0.091*	5.06	0.083*	4.62
Honduras	370	0.052	5.59	0.108*	6.71	0.086*	6.46
Hungary	311	0.057	2.85	0.068*	3.82	0.050	3.50
India	368	0.052	-0.29	0.143*	11.97	0.128*	13.51
Indonesia	368	0.052	-0.94	0.129*	11.38	0.139*	11.97
Jamaica	371	0.052	1.06	0.179*	15.07	0.144*	12.63
Kenya	372	0.052	0.16	0.029	2.19	0.023	1.99
Malawi	250	0.063	2.24	0.059	3.66	0.034	2.92
Mauritius	372	0.052	-2.31	0.078*	5.24	0.062*	4.95
Mauritania	189	0.073	-0.94	-0.073	-4.87	-0.070	-4.75
Mexico	371	0.052	0.49	0.158*	20.72	0.149*	177.5
Morocco	369	0.052	-2.72	0.087*	6.57	0.097*	8.85
Myanmar	366	0.052	2.23	0.072*	3.91	0.082*	4.49
Namibia	346	0.054	-1.56	——^{a}	——^{a}	0.023	1.99
Nepal	374	0.052	-0.55	0.107*	7.78	0.109*	7.88
Niger	369	0.052	-2.06	0.031	2.47	0.029	2.52
Nigeria	365	0.052	1.04	0.128*	6.21	0.097*	5.28
Pakistan	371	0.052	-1.93	0.051	3.41	0.057*	3.50
Paraguay	368	0.052	0.45	0.047	2.20	0.045	1.88
South Africa	371	0.052	-1.39	0.043	3.19	0.043	2.88
Sri Lanka	370	0.052	0.55	0.071*	5.72	0.082*	7.50
Suriname	295	0.059	3.72	0.177*	11.16	0.159*	12.54
Swaziland	364	0.052	-0.69	-0.071*	-4.96	-0.083*	-5.92
Turkey	242	0.064	0.92	0.062	2.14	0.133*	9.17
Venezuela	336	0.055	6.25	0.156*	10.27	0.159*	26.10
Zambia	157	0.080	-1.25	0.114*	6.84	0.145*	8.81
Zimbabwe	263	0.062	4.79	0.073*	3.29	0.066*	4.34

^{a} The null hypothesis of no bilinearity is not rejected.

The results presented in Table 3 reveal that the majority of the bilinear coefficient estimates are of a magnitude larger than the theoretically established upper limit $1/\sqrt{T}$ (see the values marked by asterisks). Although the estimation results vary with regard to the applied technique, both methods of estimation lead to similar conclusions concerning the presence of bilinearity. Bilinearity has not been established for any of the developed countries and is explosive for 26 using KF (24 using OLS) out of 40 (37) price series.

7 Conclusions

The Maximum Likelihood method has been suggested for estimating the unit root bilinear process defined by CLM. The log-likelihood function based on the Kalman filter is numerically maximized applying either the Simulated Annealing or the Newton-Raphson algorithm. We initialize the NR algorithm by the OLS estimates of the unknown parameters. If NR fails to maximize the log-likelihood function, we suggest the use of the SA algorithm.

An application to 106 price series reveals that the greater part of the significant bilinear coefficient estimates (all of them for the developing countries) are explosive.

References

[1] Byers, J.D., and D.A.Peel (1995): "Bilinear quadratic ARCH and volatility spillovers in inter-war exchange rates", *Applied Economics Letters*, 2, 215-219.

[2] Charemza, W.W., M.Lifshits, and S.Makarova (2002): "Behaviour of statistical aggregates under small bilinearity", State University of St. Petersburg, mimeo.

[3] Charemza, W.W., M.Lifshits, and S.Makarova (forthcoming): "Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results", *Journal of Economic Dynamics and Control*.

[4] Corana, A., M.Marchesi, C.Martini, and S.Ridella (1987): "Minimizing Multimodal Functions of Continuous Variables with the "Simulated Annealing" Algorithm", *ACM Transactions on Mathematical Software*, 35, 262-80.

[5] Goffe, W.L., G.D.Ferrier, and J.Rogers (1994): "Global optimization of statistical functions with simulated annealing", *Journal of Econometrics*, 60, 65-99.

[6] Granger, C.W.J., and A.P.Andersen (1978): *An introduction to bilinear time series models*. Gottingen: Vandenhoeck and Ruprecht.

[7] Hamilton, J.D. (1994): *Time Series Analysis*. Princeton: Princeton University Press.

[8] Harvey, A.C. (1989): *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

[9] Hinich, M.J., and D.M.Patterson (1985): "Evidence of Nonlinearity in Daily Stock Returns", *Journal of Business & Economic Statistics*, 3, 69-77.

[10] MacKinnon, J.G. (1991): "Critical Values for Cointegration Tests", in Engle, R.F. and C.W.J.Granger, eds., *Long Run Economic Relationships*, Oxford: Oxford University Press, 267-276.

[11] Maravall, A. (1983): "An Application of Nonlinear Time Series Forecasting", *Journal of Business & Economic Statistics*, 1, 66-74.

[12] Marsi, S.F., G.A.Bekey, and F.B.Safford (1980): "A global optimization algorithm using adaptive random search", *Applied Mathematics and Computation*, 7, 353-375.

[13] Metropolis, N., A.Rosenbluth, M.Rosenbluth, A.Teller, and E.Teller (1953): "Equation of state calculations by fast computing machines", *Journal of Chemical Physics*, 21, 1087-1090.

[14] Nelder, J.A. and R.Mead (1965): "A simplex method for function minimization", *The Computer Journal*, 7, 308-313.

[15] Peel, D., and J.Davidson (1998): "A non-linear error correction mechanism based on the bilinear model", *Economics Letters*, 58, 165-170.

[16] Scheinkman, J.A., and B.LeBaron (1989): "Nonlinear Dynamics and Stock Returns", *The Journal of Business*, 62, 311-337.

[17] Subba Rao, and M.M.Gabr (1984): *An introduction to bispectral analysis and bilinear time series models.* Series Lecture notes in Statistics, 24, New York: Springer.

[18] Terdik, G. (1999): *Bilinear stochastic models and related problems of nonlinear time series analysis: a frequency domain approach.* New York: Springer.