ISSN 1032-3813
ISBN 0 7326 1026 5

MONASH UNIVERSITY

AUSTRALIA

BAYESIAN ARBITRAGE THRESHOLD ANALYSIS

Catherine S Forbes
Guyonne R.J. Kalb
Paul Kofman

Working Paper 3/97
April 1997

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Bayesian Arbitrage Threshold Analysis

by

Catherine S. Forbes, Guyonne R.J. Kalb, and Paul Kofman*

Department of Econometrics
Monash University
Clayton, VIC 3168, Australia
Tel.+61.3.99055847 Fax.+61.3.99055474

April 3, 1997

Abstract

A Bayesian estimation procedure is developed for estimating multiple regime (multiple threshold) vector autoregressive models appropriate for deviations from financial arbitrage relationships. This approach has clear advantages over classical stepwise threshold autoregressive analysis. Whereas classical procedures first have to identify thresholds and then perform piecewise autoregressions, we simultaneously estimate threshold and autoregression parameters. To illustrate the Bayesian procedure, we estimate a no-arbitrage band within which index futures arbitrage is not profitable despite (persistent) deviations from parity.

Keywords: Bayesian VAR, error correction, non-linear threshold models, index futures arbitrage

JEL-Code: C11,G13

1. Introduction

The arbitrage argument provides a powerful tool for financial economists in deriving equilibrium conditions. It is therefore no surprise that there is an extensive literature trying to explain and/or accommodate consistent deviations from well known financial arbitrage or parity relationships. Whereas it is tempting to conclude from the empirical evidence that such deviations imply market inefficiencies, any profit potential seems to be severely obstructed by arbitrage costs. Unfortunately, some aspects of the total

* The authors thank Herman van Dijk, Gael Martin, and seminar participants at Monash University for helpful comments. Paul Kofman acknowledges research support by a Monash University Faculty Research Grant.

arbitrage cost are difficult to measure, and hence it is difficult to determine when an arbitrage opportunity truly exists. One approach would be to attempt to identify the total arbitrage cost by exploiting any measurable peculiarities in the observed deviations from equilibrium. This results in searching for threshold values away from the equilibrium condition beyond which the deviations from equilibrium are large enough to imply an arbitrage opportunity.

Regime switching models seem most appropriate for describing this arbitrage threshold idea, since they distinguish between regimes based on apparent differences in the stochastic process of the variable of interest. In the arbitrage setting, we postulate a middle regime where transaction costs prevent profitable arbitrage and two (upper and lower) outer regimes where arbitrage will occur. Tong's (1983) seminal book on threshold models introduces threshold autoregressive (TAR) models relevant to economic problems. TAR models consist of piecewise linear autoregressions. They are linear in the state space of the stochastic process of the variable of interest, but nonlinear in the time domain. Usually there is a threshold variable indicating at each point in time which regime prevails. This variable can be a function of the observable history, or an unobservable latent variable, e.g., a saturating consumption utility level; see Pole and Smith (1985). Markov Chain regime switching models typically allow for unobservables to identify the regime. Potter (1995) illustrates how these apparently distinctive models can be nested in a single index generalized multivariate autoregressive (SIGMA) model class.

One appealing special case in the SIGMA class is given by the Self Exciting TAR (SETAR) model, where the regime indicator is given by the (lagged) variable of interest itself. This facilitates estimation in comparison with the unobservable indicator model. There is also a pragmatic argument in favour of using this feature. If a strictly homogeneous asset is traded simultaneously in two markets, it should have the same price in both markets. However, consider the fact that these markets are geographically separated. It will then take some time for participants in both markets to observe occasional price differences. The SETAR model will then use the (appropriately) lagged price difference as an indicator variable triggering arbitrage if this difference is sufficiently large (i.e., exceeds some threshold) to offset the transportation costs of moving the asset from the lower price to the higher price market.

A number of papers have recently appeared providing estimation methodologies to cope with threshold estimation problems. Relevant to our purposes, Balke and Fomby (1993), Yadav, Pope, and Paudyal (1994), and Martens, Kofman, and Vorst (1996) propose the SETAR modelling procedure to identify arbitrage thresholds or triggers. Unlike other threshold papers this implies that there is a well defined economic model underlying the analysis. The underlying threshold autoregression methodology used, based on Tsay (1989), relies on recursive arranged autoregressions. Visual inspection of the residual scatterplots of the arranged regressions leads to an indication of a range of possible thresholds. A minimizing sum of squared errors grid search on this range then defines the threshold. In practice, it proves difficult to operationalize this technique. Whereas the method seems to be rather powerful in detecting evidence of non-linearity, its threshold estimation power is much less impressive. The problems encountered in deriving suitable threshold values are rather evident in for e.g., Martens, Kofman, and Vorst (1996). Appropriately picking the grid range from scatterplots is tedious at best. In particular, the ordering of the deviations (ascending or descending) influences the range of possible thresholds and the probability of detection. Due to the very concept of arbitrage, any crossing of the thresholds will by definition only persist for a very short time. The subsequent small sample size for the outer regimes hampers proper threshold identification. Balke and Fomby (1993) suggest that the TAR modelling procedure may then not be useful.

An obvious alternative is then to look at a Bayesian threshold methodology. A few papers have recently appeared[1], discussing univariate TAR models from a Bayesian perspective. Geweke and Terui (1993) develop a 2-regime AR-model and provide analytical solutions for the marginal posterior density of the threshold variable. Koop (1996) considers a similar model, but focuses on subsequent impulse response analysis. We propose a Bayesian estimation methodology for a threshold vector autoregression model which simultaneously estimates thresholds, delay, and

[1] Pole and Smith (1985) present a Bayesian analysis of the related threshold regime switching models.

interaction parameters. Thus, it avoids some of the troublesome aspects of the classical methodology predominantly used in this line of research. We extend Geweke and Terui by considering multiple thresholds in a vector autoregressive (VAR) setting. We illustrate that it is possible to obtain a straightforward numerical solution for this case. In addition, we use 'Rao-Blackwellized' estimators for calculating marginal posterior summaries. Chen and Lee (1995) suggest that a simple alternative to Geweke and Terui's model can be obtained via Markov chain Monte Carlo (MCMC) methods. We argue that our solution dominates this simulation approach when noninformative prior distributions are used.

The remainder of this paper is organized as follows. Section 2 introduces a threshold error correction model for arbitrage in a simple context and then extends the ideas to a more general arbitrage case. Section 3 gives an outline of the computational method proposed for obtaining an independent sample from the joint posterior distribution, and proceeds in providing details regarding the calculation of marginal posterior quantities of interest, such as posterior moments and marginal densities. We then apply our model empirically to index futures arbitrage in the S&P500 stock index. Details of the derivations relevant to Section 3 are given in an appendix.

2. A Threshold Error Correction Model for Arbitrage

Threshold autoregressive models have been applied to economic problems. Unfortunately, for some applications the rationale for a threshold or the interpretation of the autoregressive estimates is difficult to grasp. One area where we might expect meaningful and interpretable applications is given by financial arbitrage relations. Our starting point, therefore, is the simplest of these; the law of one price (LOOP). According to the LOOP, a strictly identical asset which is traded in separate markets should have the same price at the same time to avoid arbitrageurs from buying in the cheapest market and simultaneously selling in the dearest market. We can specify,

$$z_t = P_t - p_t, \tag{1}$$

where P_t is the price of the asset at time t in market 1, and p_t is the price of the same asset at time t in market 2. While it is possible for the price wedge z_t to be occasionally non-zero, it should have zero expectation and display strong mean-reverting behaviour. Furthermore, if the z_t are sufficiently small, the costs of arbitrage will not outweigh the profits. This implies that it is possible for z_t to be persistently positive or negative. Thus, we use (1) to construct the hypothesized error correction terms, which will become 'active' as soon as some arbitrage preventing threshold is surpassed. Once we are able to identify this threshold, we can split up our sample into regimes; *outer* regimes where arbitrage is effectively driving asset prices back towards their no-arbitrage value, and an *inner* regime where asset prices are (relatively) free to diverge from this value, i.e., behave locally as a random walk.

To describe the threshold error correction model (ECM), we begin with some notation. Define three regimes $R^{(j)}$, j = 1, 2 and 3, by $R^{(j)} = \{z_t : r_{j-1} \leq z_{t-d} < r_j\}$, with $r_0 = -\infty$ and $r_R = +\infty$. Here d is known as the *threshold lag* or *delay parameter*, and the vector $r' = (r_1, r_2)$ gives the threshold values which determine the regimes. Let q be an index for the observed data such that $z_q \leq z_{q+1}$ for all q = *1*, ..., *T-1*. Using this ordering, there correspond values of $q = Q^{(j)}$ where $z_{Q^{(j)}} \in R^{(j)}$ and $z_{Q^{(j)}+1} \in R^{(j+1)}$, for j = *1*, and *2*, and $Q^{(3)} = T - Q^{(1)} - Q^{(2)}$.

The ECM describes the distribution of the changes in the observed prices, given the values of the unknown parameters, past changes in prices and deviations from arbitrage equilibrium. We have, for each regime where $z_{q-d} \in R^{(j)}$,

$$\begin{aligned} \Delta P_q &= c_P^{(j)} + \sum_{k=1}^{L_1^{(j)}} \phi_{P,k}^{(j)} \Delta P_{q-k} + \sum_{k=1}^{L_2^{(j)}} \theta_{P,k}^{(j)} \Delta p_{q-k} + \gamma_P^{(j)} z_{q-d} + \varepsilon_{P,q}^{(j)} \\ \Delta p_q &= c_p^{(j)} + \sum_{k=1}^{L_3^{(j)}} \phi_{p,k}^{(j)} \Delta P_{q-k} + \sum_{k=1}^{L_4^{(j)}} \theta_{p,k}^{(j)} \Delta p_{q-k} + \gamma_p^{(j)} z_{q-d} + \varepsilon_{p,q}^{(j)} \end{aligned} \qquad (2)$$

where ϕ and θ are the autoregressive ('short-run') parameters and γ is the equilibrium adjustment ('long-run') parameter. Now, whenever z_{q-d} does not exceed either the upper or lower threshold, γ_P and γ_p are zero and prices P and p do not respond to disequilibria in the LOOP. Whenever z_{q-d} does exceed either threshold, γ_P and/or γ_p are non-zero and negative, respectively positive, to restore equilibrium. The constants

c, the drift in price changes, may also contribute to restore equilibrium if they have the appropriate (opposing) signs. Conditionally on regime $R^{(j)}$ and the other parameters, we assume that the random vectors

$$\varepsilon_q^{(j)} = \begin{pmatrix} \varepsilon_{P,q}^{(j)} \\ \varepsilon_{p,q}^{(j)} \end{pmatrix} \sim \text{independent Normal } \left(0_2, \Sigma^{(j)}\right) \tag{3}$$

within each regime $R^{(j)}$. The above ECM for each regime can then be written as a vector autoregression as follows:

$$y_q = c^{(j)} + A_1^{(j)} y_{q-1} + \ldots + A_L^{(j)} y_{q-L} + \gamma^{(j)} z_{q-d} + \varepsilon_q^{(j)} \tag{4}$$

where

$$y_q = \begin{pmatrix} \Delta P_q \\ \Delta p_q \end{pmatrix}$$

$$c^{(j)} = \begin{pmatrix} c_P^{(j)} \\ c_p^{(j)} \end{pmatrix}$$

$$A_i^{(j)} = \begin{bmatrix} \phi_{P,i}^{(j)} & \theta_{P,i}^{(j)} \\ \phi_{p,i}^{(j)} & \theta_{p,i}^{(j)} \end{bmatrix} \tag{5}$$

$$\gamma^{(j)} = \begin{pmatrix} \gamma_P^{(j)} \\ \gamma_p^{(j)} \end{pmatrix}$$

$$\varepsilon_{q-d}^{(j)} = \begin{pmatrix} \varepsilon_{P,q-d}^{(j)} \\ \varepsilon_{p,q-d}^{(j)} \end{pmatrix}$$

for $q = Q^{(j-1)} + 1, \ldots, Q^{(j)}$ with $Q^{(0)} = 0$ and $L^{(j)} = \max\left\{L_1^{(j)}, L_2^{(j)}, L_3^{(j)}, L_4^{(j)}\right\}$.

2.1 A General Arbitrage Threshold ECM

The equilibrium relationship implied by arbitrage can be more general than what is indicated in (1), in that more than two variables may be involved. For the general arbitrage threshold ECM, we define the univariate mispricing error variable

$$z_t = F_t(y_t) \tag{6}$$

where y_t is an $m \times 1$ vector of prices comprising the equilibrium and the functions $F_t : \mathcal{R}^m \to \mathcal{R}^1$ are assumed to be given by the arbitrage theory in this context. In the LOOP case, $y_t' = (\Delta P_t, \Delta p_t)$ and $F_t(y) = (1, -1)y$.

We can write the VAR model for each regime $R^{(j)}, j = 1, 2, 3$ in compact form

$$Y^{(j)} = X^{(j)}B^{(j)} + U^{(j)}, \tag{7}$$

where

$$Y^{(j)'} := \left[y_{Q^{(j-1)}+1}, \ldots, y_{Q^{(j)}} \right],$$

$$X^{(j)'} := \left[X_{Q^{(j-1)}+1}, \ldots, X_{Q^{(j)}} \right], \text{ where}$$

$$X_q := \begin{pmatrix} 1 \\ y_q \\ \vdots \\ y_{q-L+1} \\ z_{q-d} \end{pmatrix} \tag{8}$$

$$B^{(j)'} := \left[c^{(j)}, A_1^{(j)}, \ldots, A_L^{(j)}, \gamma^{(j)} \right], \text{ and}$$

$$U^{(j)'} := \left[\varepsilon_{Q^{(j-1)}+1}, \ldots, \varepsilon_{Q^{(j)}} \right].$$

Note that $Y^{(j)}$ is a $N^{(j)}$ *x m* matrix of observed variables, where *m* is the number of equations in the ECM and $N^{(j)}$ is the number of observations in regime $R^{(j)}$; $X^{(j)}$ is an $N^{(j)}$ *x* $k^{(j)}$ matrix of regressors, where $k^{(j)}=m(L^{(j)}+2)$ and $L^{(j)}$ is the maximum number of lags included in the ECM; $B^{(j)}$ is a $k^{(j)}$ *x m* matrix of regression coefficients and $U^{(j)}$ is an $N^{(j)}$ *x m* matrix of errors for regime $R^{(j)}$. So far we have assumed the regimes to be fixed and the values of *r* and *d* known. However, it is one of our objectives to estimate these parameters, and hence we treat them as unknown parameters. Note that variables indexed by *j* are implicitly functions of both *r* and *d*. We proceed for a general *m*-dimensional ECM and apply it to intraday futures and index returns data (where *m*=2.) in Section 4.

Martens, Kofman, and Vorst (1996) estimate an arranged autoregression in z_q based on Tsay's Threshold Autoregressive (TAR) method, which can be used to define the three regimes. Once the regimes are defined, an ECM is then estimated *for each regime,* for each equation in (2). This two-step procedure, while intuitively appealing for describing threshold concepts, does not by its very nature estimate parameters *simultaneously.* In addition, the two step procedure does not utilize a single representation of the dynamic structure implicit in the setting, and consequently the assumptions of the TAR model and the ECM are potentially conflicting.

We therefore prefer to use a Bayesian approach for analyzing the ECM, while imposing a threshold component directly to capture the desired feature of differing behaviour of the z_t variables when they are out of equilibrium. Balke and Fomby (1993) also discuss using thresholds in error correction representations, however they do not actually use the Bayesian approach in their analyses. One of the benefits of the our approach is that we are able to obtain simultaneous estimation of all parameters considered. In addition, we obtain posterior density estimates for both single parameters and various combinations of parameters of particular interest. That is, we can obtain the range of plausible parameter values and, given the data, the probabilities associated with these values.

3. Bayesian Threshold ECM Estimation

Given the threshold and delay parameters, r and d, respectively, Bayesian analysis of the ECM in each regime follows a standard Bayesian analysis for vector autoregressive (VAR) processes. See Lütkepohl (1993) and Zellner (1971). If we assume for the moment that the values of r and d are known, we have then the following results for the posterior distributions of the autoregression coefficients $b^{(j)}$ = vec($B^{(j)}$) and variance-covariance matrices $\Sigma^{(j)}$, for $j = 1, 2, 3$ under the assumption of the following prior for $b^{(j)}$ and $\Sigma^{(j)}$,

$$p(b^{(j)},\Sigma^{(j)}|r,d) \propto \prod_{j=1}^{3}\left|\Sigma^{(j)}\right|^{-(m+1)/2}. \tag{9}$$

We have

$$\Sigma^{(j)} \text{ given } Y,r,d \sim \text{ independent } \mathrm{IW}_m\left(v^{(j)},S^{(j)}\right) \tag{10}$$

$$b^{(j)} \text{ given } \Sigma^{(j)},Y,r,d \sim \text{ independent } N_{mk^{(j)}}\left(\hat{b}^{(j)},\Sigma^{(j)} \otimes X^{(j)\prime} X^{(j)}\right) \tag{11}$$

where $\hat{b}^{(j)} = vec(\hat{B}^{(j)})$ and $\hat{B}^{(j)} = \left(X^{(j)\prime} X^{(j)}\right)^{-1} X^{(j)\prime} Y^{(j)}$ is the OLS estimator of $B^{(j)}$ given r and d, $v^{(j)} = N^{(j)} - k^{(j)}$ and $S^{(j)} = \left(Y^{(j)} - X^{(j)}\hat{B}^{(j)}\right)^{\prime}\left(Y^{(j)} - X^{(j)}\hat{B}^{(j)}\right)$. Here $\mathrm{IW}_m(v,S)$ refers to an inverted Wishart distribution for an m x m positive definite symmetric matrix with degrees of freedom parameter v and matrix parameter S.

Since r and d are unknown, to produce a sample from the joint posterior distribution of $\left\{b^{(j)}, \Sigma^{(j)}, j=1,2,3\right\}, r$ and d, we must first be able to obtain r and d from the marginal posterior distribution $p(r,d|y)$. Once we have r and d, we can sample $\left\{\Sigma^{(j)}, j=1,2,3\right\}$ and then $\left\{b^{(j)}, j=1,2,3\right\}$ using (10) and (11), respectively.

Following Geweke and Terui (1993), we assume a uniform prior $p(r,d) \propto 1$, for $a_L \le r_1 < r_2 \le a_U$ and $d = 1, 2, \ldots, D$, and derive an expression for the marginal posterior distribution for r and d

$$p(r,d|y) \propto \prod_{j=1}^{R} \pi^{-m\nu^{(j)}/2} \prod_{j=1}^{m} \Gamma\left[\left(\nu^{(j)}+1-i\right)/2\right] \left|S^{(j)}\right|^{-\nu^{(j)}/2} \left|X^{(j)\prime} X^{(j)}\right|^{-m/2}. \tag{12}$$

Details of derivations of (10) (11) and (12) are given in the appendix. Notice that potentially many values for the threshold and delay parameters are included in the marginal posterior distribution of r and d. The values that appear most likely, given the observed data y, will have the greatest influence on the estimates of $\Sigma^{(j)}$ and $b^{(j)}$.

To actually sample from the above joint posterior $p(r,d|y)$, we numerically normalize the right hand side of (12) and sample from the resulting empirical distribution. Then, given sampled values of r and d, the $\Sigma^{(j)}$ are sampled from independent $IW_m(\nu^{(j)}, S^{(j)})$ distributions. Finally, given r, d and $\Sigma^{(j)}$, the elements of $b^{(j)} = vec(B^{(j)})$ are jointly normally distributed with mean values given by the OLS estimator of $b^{(j)}$, namely $\hat{b}^{(j)} = vec(\hat{B}^{(j)})$, and variance-covariance matrix of $b^{(j)}$ given by $\left(\Sigma^{(j)} \otimes X^{(j)\prime} X^{(j)}\right)$. In this manner, a sample of any size can be obtained from the full joint posterior distribution.

3.1 Estimating Marginal Posterior Distributions

The above method describes how to sample from the full joint posterior distribution of r,d and $\Sigma^{(j)}$ and $b^{(j)}$ for each regime. However, often we are interested in marginal posterior distributions for each of the variables, and specifically marginal posterior means and variances for each. One approach is to use Monte Carlo sample summaries for the marginal posterior means and variances for each variable based on the values obtained via Monte Carlo sampling of $r, d, \Sigma^{(j)}$ and $b^{(j)}$, for $j = 1, 2, \ldots, R$. However, the so-called 'Rao-Blackwellized' estimators of marginal densities and moments will provide Monte Carlo estimators with smaller (Monte Carlo) mean squared error. See Gelfand and Smith (1990). Here we report on the form of these estimators for marginal posterior means, variances and densities for the regression coefficients in each regime.

Let $b_l^{(j)}$ denote the l^{th} element of the regression coefficient vector $b^{(j)}$ in the j^{th} ECM and $\left(r_i, d_i, \Sigma_i^{(j)}\right)$ for $i = 1,2,\ldots,M$ be a sample of size M from the posterior distribution. A 'Rao-Blackwellized' estimate of the posterior mean of $b_l^{(j)}$ is given by a sample average of the OLS estimates $\hat{b}_{l,i}^{(j)}$ of $b_l^{(j)}$

$$\hat{E}_{rb}[b_l^{(j)} | y] = \frac{1}{M}\sum_{i=1}^{M}\hat{b}_{l,i}^{(j)}. \tag{13}$$

A Rao-Blackwellized estimate of the marginal posterior variance of $b_l^{(j)}$ can be similarly obtained using

$$V\hat{a}r_{rb}\left(b_l^{(j)2} | y\right) = \left[\frac{1}{M}\sum_{i=1}^{M}\left\{V_{l,i}^{(j)} + \left(\hat{b}_{l,i}^{(j)}\right)^2\right\}\right] - \left(\hat{E}_{rb}\left[b_l^{(j)} | y\right]\right)^2 \tag{14}$$

where $V_{l,i}^{(j)}$ is the l^{th} diagonal element of $\left(\Sigma_i^{(j)} \otimes X_i^{(j)\prime} X_i^{(j)}\right)$. Covariances between elements of $b^{(j)}$ can also be obtained.

Rao-Blackwellized estimates of marginal posterior density functions are also possible by averaging the marginal (conditional) posterior densities of $b_l^{(j)}$, given sampled values $\left(r_i, d_i, \Sigma_i^{(j)}\right)$, over a grid of possible values. That is, for each point b, we have

$$\hat{p}_{rb}(b | y) = \frac{1}{M}\sum_{i=1}^{M}(2\pi)^{-1/2} |V_{l,i}^{(j)}|^{-1/2} \exp\left\{\frac{-1}{2V_{l,i}^{(j)}}\left(b - \hat{b}_{l,i}^{(j)}\right)^2\right\}. \tag{15}$$

Bivariate density estimates can be similarly obtained.

In summary, due to the particular choice of the prior distribution, the marginal posterior density for r and d is available analytically up to a normalizing constant. Sampling from this distribution is straightforward. Conditional on sampled pairs r and d, simulation of the remaining variables from the full joint posterior distribution is straightforward, and as a result various summary statistics of the posterior distribution are readily accessible, to any desired degree of accuracy.

We have given a detailed approach for analyzing a threshold error correction model under the assumption that there are two threshold resulting in three regimes. Extension to any fixed number of regimes is straightforward. In addition, our approach may be appropriate for some alternative specifications of the prior distribution, however each choice would need to be evaluated individually to determine if direct sampling from the joint posterior distribution is possible. In particular, if a different prior distribution was desired and was of a form that analytical reduction to the marginal posterior density of r and d was not possible, other techniques, such as using Markov chain Monte Carlo (MCMC) methods might be more suitable. For a discussion of a Bayesian analysis of a univariate TAR model using MCMC methods, refer to Chen and Lee (1995). However, we believe unless formal prior information is available that is not compatible with our specification, the approach presented here is preferable in that it does not require more elaborate computational techniques.

4. Empirical Application

In this section, we apply our estimation procedure to index futures arbitrage. This type of arbitrage consists of simultaneously buying (short-selling) the stock index and selling (buying) the stock index futures whenever their prices diverge by more than the cost of 'carrying' the stock index through time until maturity of the futures contract. Brenner and Kroner (1995) give the following cost-of-carry expression,

$$F_{t,T} - S_t = \left(r_{t,T} - q_{t,T}\right)\left(T - t\right) + z_t \ . \qquad (16)$$

where F_t, S_t are the logarithms of respectively futures and stock index prices, $r_{t,T}$ is the risk-free interest rate, $q_{t,T}$ is the dividend yield on the stock index, and $(T\text{-}t)$ is the time to maturity of the futures contract. For arbitrage to be profitable, we require that z_t has to exceed a certain (absolute) value determined by transaction costs. However, this will not be a sufficient determinant for the arbitrage 'wedge', since we consider an intertemporal arbitrage relation. Hence, we distinguish between two types of no-arbitrage costs. The 'direct costs', such as transaction costs, short-selling restrictions and unequal borrowing and lending rates are relatively easy to measure. The 'indirect cost' components, such as index tracking error, execution risk, and dividend and interest rate risk on the other hand are much more difficult to determine, necessitating estimation of the thresholds to determine the total no-arbitrage band .

For illustration we use one month (May 1995) of intraday transaction data for the S&P500 stock index and its (nearest delivery) futures contract traded at the Chicago Mercantile Exchange. The maturity date for this particular contract is June. Even though theoretically spot and futures prices converge towards maturity, it is well known empirically that this does not affect the stochastic behaviour of z_t, the mispricing error, well in advance of maturity. To avoid (or moderate) typical microstructural problems like bid-ask bounce and infrequent trading, we construct a one minute bivariate price series, generating 7,060 observations.

For detailed information on the univariate statistical properties of these series we refer to Martens, Kofman, and Vorst (1996). Most important for our purpose is the fact that futures and spot price are cointegrated, where the appropriate cointegrating relationship is given by (16) above. Instead of focusing on this cointegration relationship, we straightforwardly specify (and estimate) our implied threshold vector error correction model,

$$\begin{aligned} \Delta \ln F_{\pi_t,T} &= c_F^{(j)} + \sum_{k=1}^{L_1^{(j)}} \phi_{F,k}^{(j)} \Delta \ln F_{\pi_t-k,T} + \sum_{k=1}^{L_2^{(j)}} \theta_{F,k}^{(j)} \Delta \ln S_{\pi_t-k} + \gamma_F^{(j)} z_{\pi_t-d} + \varepsilon_{F,\pi_t}^{(j)} \\ \Delta \ln S_{\pi_t} &= c_S^{(j)} + \sum_{k=1}^{L_3^{(j)}} \phi_{S,k}^{(j)} \Delta \ln F_{\pi_t-k,T} + \sum_{k=1}^{L_4^{(j)}} \theta_{S,k}^{(j)} \Delta \ln S_{\pi_t-k} + \gamma_S^{(j)} z_{\pi_t-d} + \varepsilon_{S,\pi_t}^{(j)} \end{aligned} \quad (17)$$

where z_t is as defined in (16). We postulate a three-regime model for our analysis. Yadav, Pope, and Paudyal (1994) and Martens, Kofman, and Vorst (1996) allow for even larger models. Yadav, Pope, and Paudyal, for example, argue that at different thresholds different arbitrageurs become active depending on their relative cost (dis)advantage. This is especially the case if arbitrage capital is constrained. It seems unlikely to us that large sophisticated financial institutions will leave any opportunities to less equipped arbitrageurs. Fully automated program trading, for example, automatically triggers the appropriate arbitrage strategy based on preset mispricing thresholds. It is also difficult to imagine a shortage of arbitrage capital for these least-cost arbitrageurs. For the purpose and transparency of our analysis, we therefore prefer to focus on a small model. Larger models with many more regimes can nevertheless be accommodated for in this methodology.

A second choice we have to make is the lag length $L_i^{(j)}$ for each regime. Once more, we take a pragmatic view. For all regimes we set the lag length equal to eight, given the results in Martens, Kofman, and Vorst (1996). We realize the risk of a potentially overparameterized model, but for now want to focus on estimation of the other parameters of interest, i.e., d, r, ϕ, θ, and γ, conditionally on this lag length choice.

First, we report on the posterior density of d where we find overwhelming evidence for a threshold lag of just one minute. We do not observe any probability at higher lags (we restrict the analysis to a maximum threshold lag of four minutes[2]). Next, Table 1 provides the marginal posterior probabilities for some individual threshold candidates pairs (r_1, r_2). As the six pairs of threshold values given account for approximately 98% of the marginal posterior probability[3], we give only these values here. No other pair contributes more than 0.4% of the total probability. The lower threshold has a modal value of -0.10381 percent, the upper threshold has a modal value of 0.12763 percent, and hence they appear to be close to being symmetric around zero. Translated into commonly used arbitrage index points, this amounts to a no-arbitrage band of 1.03 index points.[4] This value corresponds to the band with the greatest posterior probability (approximately 88%). We find some evidence of alternative threshold candidates, e.g., -0.09542 percent for the lower bound and 0.12763 percent for the upper threshold. According to this data, the smallest no-arbitrage band we observe with any significant probability is 0.99 index points, corresponding to approximately 10% of the posterior probability. Martens, Kofman, and Vorst (1996) find a no-arbitrage band of 1.61 index points for the same sample period. The 'direct' costs involved in an arbitrage transaction consist mainly of bid-ask spread costs for the least-cost

[2] Program traders are guaranteed execution of trades within 3 minutes.
[3] Empirical probability associated with numerical approximation to marginal posterior distribution of r.
[4] We find this value by multiplying the threshold bandwidth and the average cash index level for May, which is 445.25.

arbitrageurs. Given that this cost component is about 0.75 index points for our sample period, we could argue that a value between 1.0 and 1.03 index points is a more likely candidate for the total cost threshold.

Table 1. Bayesian Threshold Candidates

Threshold values		Posterior	Index point
r_1	r_2	probability	no arbitrage band
-0.10381	0.12627	0.019	1.03
-0.10381	0.12695	0.099	1.03
-0.10381	0.12763	0.707	1.03
-0.10381	0.12831	0.058	1.03
-0.09542	0.12695	0.012	0.99
-0.09542	0.12763	0.084	0.99
	total	0.979	

Having discussed d and r, we can now move on to our error correction model estimates for each regime. We reviewed the marginal posterior density estimates for each individual regression coefficient in the ECM. For our data, all of the plots appeared to match a normal distribution with mean and standard error given by the estimates in Table 2. Therefore, we do not present them all here. A typical finding for index-futures arbitrage is that the futures market tends to be more informationally efficient than the cash market. In lead-lag terms, this implies that futures returns lead the stock index returns. Kawaller, Koch, and Koch (1987), e.g., find that for 1984-1985 the S&P futures price significantly leads the spot price up to 20-45 minutes. Sometimes the spot price is found to lead, but this lead seldom extends beyond one minute. We find a similar lead-lag structure though less pronounced for the futures

Table 2. Bayesian T-VAR estimates *(Posterior standard deviations)* [a]

	Lower Regime		Central Regime		Upper Regime	
	ΔF	ΔS	ΔF	ΔS	ΔF	ΔS
C	0.00005	0.00015	0.00000	0.00000	0.00021	-0.00011
	(0.00009)	*(0.00006)*	*(0.00000)*	*(0.00000)*	*(0.00011)*	*(0.00005)*
ϕ_1	-0.12447	0.15643	-0.05627	0.04083	0.05203	0.04588
	(0.05566)	*(0.03718)*	*(0.01490)*	*(0.00625)*	*(0.06240)*	*(0.03104)*
ϕ_2	-0.03962	0.29075	0.01948	0.10659	-0.00762	0.17707
	(0.05763)	*(0.03704)*	*(0.01473)*	*(0.00623)*	*(0.06287)*	*(0.03124)*
ϕ_3	-0.01052	0.17205	0.03519	0.10486	0.01773	0.11109
	(0.06345)	*(0.04045)*	*(0.01491)*	*(0.00626)*	*(0.06775)*	*(0.03363)*
ϕ_4	0.06107	0.12165	0.01021	0.08488	-0.01979	0.05400
	(0.06648)	*(0.04222)*	*(0.01506)*	*(0.00632)*	*(0.06675)*	*(0.03318)*
ϕ_5	0.02123	0.13770	0.01200	0.06408	0.02828	-0.00974
	(0.06370)	*(0.04102)*	*(0.01507)*	*(0.00634)*	*(0.06991)*	*(0.03473)*
ϕ_6	-0.07004	0.09002	-0.00861	0.04895	0.04478	-0.00998
	(0.06903)	*(0.04473)*	*(0.01485)*	*(0.00628)*	*(0.06857)*	*(0.03409)*
ϕ_7	-0.00048	0.09363	0.00085	0.04042	0.12125	0.02603
	(0.07183)	*(0.04561)*	*(0.01439)*	*(0.00604)*	*(0.06876)*	*(0.03414)*
ϕ_8	-0.11277	0.05900	-0.01073	0.01998	0.00793	-0.01393
	(0.06583)	*(0.04163)*	*(0.01389)*	*(0.00584)*	*(0.06453)*	*(0.03207)*
θ_1	0.08200	0.07006	-0.05693	-0.04329	0.10122	0.23187
	(0.10423)	*(0.06715)*	*(0.02913)*	*(0.01223)*	*(0.10663)*	*(0.05299)*
θ_2	-0.18371	0.04821	-0.04360	-0.01548	-0.02401	-0.04405
	(0.11368)	*(0.07261)*	*(0.02836)*	*(0.01192)*	*(0.11388)*	*(0.05659)*
θ_3	-0.05454	-0.21135	0.02111	-0.01344	-0.19458	-0.04320
	(0.12602)	*(0.08057)*	*(0.02782)*	*(0.01169)*	*(0.11623)*	*(0.05778)*
θ_4	0.14164	0.17854	0.00794	-0.00224	-0.00093	0.01140
	(0.11573)	*(0.07436)*	*(0.02776)*	*(0.01166)*	*(0.11723)*	*(0.05824)*
θ_5	0.42977	-0.11556	0.01593	0.02194	-0.06630	-0.01427
	(0.12369)	*(0.07867)*	*(0.02744)*	*(0.01153)*	*(0.10813)*	*(0.05373)*
θ_6	0.07636	0.04406	0.03863	0.00159	0.07089	0.07668
	(0.10947)	*(0.07005)*	*(0.02684)*	*(0.01130)*	*(0.11177)*	*(0.05554)*
θ_7	-0.22189	-0.04924	-0.00389	0.01807	-0.24684	0.01205
	(0.10819)	*(0.06847)*	*(0.02589)*	*(0.01086)*	*(0.11102)*	*(0.05518)*
θ_8	-0.06405	0.03552	0.00655	0.01732	0.18237	0.03159
	(0.11379)	*(0.07227)*	*(0.02543)*	*(0.01068)*	*(0.10438)*	*(0.05189)*
γ	0.00056	0.00132	0.00002	0.00017	-0.00130	0.00110
	(0.00073)	*(0.00047)*	*(0.00008)*	*(0.00003)*	*(0.00076)*	*(0.00038)*

$$\Sigma_L = \begin{bmatrix} 0.0000000975 & 0.0000000170 \\ 0.0000000170 & 0.0000000397 \end{bmatrix}, \Sigma_C = \begin{bmatrix} 0.0000000793 & 0.0000000085 \\ 0.0000000085 & 0.0000000140 \end{bmatrix}, \Sigma_U = \begin{bmatrix} 0.0000001505 & 0.0000000260 \\ 0.0000000260 & 0.0000000372 \end{bmatrix}$$

[a] Rao-Blackwellized means and standard errors.

lead (up to eight minutes).[5] This is most apparent in the upper regime where we only observe a significant two- and three-minute lead of the futures returns.

The parameter $\phi_{F,1}$ is significantly negative for the lower and middle regimes and ambiguously so for the upper regime. This illustrates the bid-ask spread induced bounce, which implies significant negative first-order serial correlation in futures price changes. Parameter $\theta_{S,1}$ on the other hand, is only significantly negative in the middle regime. This can be an indication of the fact that spot price changes in the outer regimes are mostly driven by infrequent trading instead of the bid-ask bounce. The positive serial correlation implied by infrequent trading more than offsets the bid-ask bounce driven negative serial correlation. This phenomenon also explains why most of the equilibrium adjustment seems to occur in the spot (stock index) price changes.

For illustration, the Rao-Blackwellized marginal posterior density estimates for the first-order 'cross-autoregression' parameters for each regime are shown in Figure 1. The parameter $\theta_{F,1}$ appears to be only significantly different from zero in the middle regime, while parameter $\phi_{S,1}$ is significant for all regimes (strongest in the outer regimes). This reinforces our finding that the spot market is responding to futures market information, but much less so the other way around.

Summarizing the short-run dynamics, it is quite apparent that spot price changes are more time- and cross-dependent than futures price changes. Also apparent is the normality of the posterior plots for those parameters. The near normality appears to be due to the sharpness of the marginal posterior distribution for r and d and the large sample size.

Somewhat puzzling is the significant drift term for the index returns in the outer regimes. For both lower and upper regimes, the signs imply a drift away from the no-arbitrage band instead of a reversion. However, it is not so easy to interpret individual parameters in this model. Apparently, the drift is more than offset by the lagged variables and error correction parameter. For a closer investigation of the error

[5] Note that we have restricted our lead/lag order to eight minutes. Martens, Kofman and Vorst (1996) do find some longer significant lags.

correction parameters we plot the joint posterior density of γ_F and γ_S in Figures 2 for each regime.

For strong error correction behaviour, we expect to observe a predominance of probability in the second quadrant (negative values for γ_F and positive values for γ_S). This is most evident in the upper regime, and least so in the middle regime. Broadly speaking, this confirms our arbitrage threshold model. Once again, the significant positive γ_s confirms that it is the stock market restoring equilibrium. In the outer regimes this is evidently stronger, most likely due to arbitrage.

5. Conclusion

Proper identification of regimes in a nonlinear threshold VAR is notoriously difficult. In this paper we develop a Bayesian approach which is better suited to the intricate 'endogenous' nature of the problem at hand. We simultaneously estimate the delay parameter, threshold values, and VAR parameters. Thus, our model avoids the somewhat awkward stepwise classical approach.

We also show that it is possible to obtain analytical solutions in a multiple threshold model. Chen and Lee's (1995) claim that their Gibbs sampling procedure outperforms the analytical solution in Geweke and Terui (1993) does not seem to be valid for the problem considered in this paper. Whereas the Gibbs sampling approach may be worthwhile for the more complicated case of incorporating informative prior information, the analytical solution seems preferable in our case.

A number of issues have yet to be resolved. First, we 'assume' a cointegrating relationship and focus straightforwardly on the error correction representation. Whereas this seems to be a valid exercise for parity relationships imposed by economic theory, this is often not the case for other interesting economic applications where we do not have a solid economic foundation. Second, a proper analysis would incorporate a model selection strategy. Selection of an appropriate lag structure could be achieved by applying the PIC (Posterior Information Criterion) procedure, see Phillips (1996). PIC can be used to achieve joint order selection of the cointegrating rank and selection of the autoregressive lag length. This may, however, be more complicated in our interdependent threshold setting.

An interesting extension will be given by impulse response functions (IRF's) analysis as in Koop (1996), for our threshold VAR model. However, this is already rather complicated in a two-regime model, let alone a multiple regime model. The explosion of possible histories and future trajectories might make a meaningful IRF analysis rather intractable. Another methodological extension can be found in the smooth transition literature. Our thresholds create a discontinuity, whereas one might argue that many economic regime changes occur in a rather smooth way. The next step will therefore inevitably include a Bayesian STAR model.

A second extension is given by Geweke and Terui who exploit their model to predict regime changes. We might be similarly interested in the probability that a future price variable will lie, for example, in the upper arbitrage regime. Conceptually, this should be rather easy to achieve.

References

Balke, N.S., and T.B. Fomby, 'Threshold Cointegration,' 1997, *International Economic Review*, forthcoming.

Brenner, R.J., and K.F. Kroner, 'Arbitrage, Cointegration, and Testing the Unbiasedness Hypothesis in Financial Markets,' *Journal of Financial and Quantitative Analysis*, 1995, 30, 23-42.

Chen, C.W.S., and J.C. Lee, 'Bayesian Inference of Threshold Autoregressive Models,' *Journal of Time Series Analysis*, 1995, 16, 483-492.

Gelfand, A.E and A.F.M. Smith, 'Sampling-Based Approaches to Calculating Marginal Densities,' *Journal of the American Statistical Association*, 1990, 85, 410, 398-409.

Geweke, J., and N. Terui, 'Bayesian Threshold Autoregressive Models for Nonlinear Time Series, *Journal of Time Series Analysis*, 1993, 14, 441-454.

Koop, G., 'Parameter Uncertainty and Impulse Response Analysis,' *Journal of Econometrics*, 1996, 72, 135-149.

Lütkepohl, H. *Introduction to Multiple Time Series Analysis*, 1993, 2nd edition, Springer, New York.

Martens, M., P. Kofman, and T.C.F. Vorst, 'A Threshold Error Correction Model for Intraday Futures and Index Returns,' *Research Symposium Proceedings of the Chicago Board of Trade*, 1996, Summer, 231-265.

Phillips, P.C.B., 'Econometric Model Determination,' *Econometrica*, 1996, 64, 763-812.

Phillips, P.C.B., 'To criticize the critics: An objective Bayesian analysis of stochastic trends,' *Journal of Applied Econometrics*, 1991, 6, 333-364.

Pole, A.M., and A.F.M. Smith, 'A Bayesian Analysis of Some Threshold Switching Models,' *Journal of Econometrics*, 1985, 29, 97-119.

Potter, S.M., 'A Nonlinear Approach to US GNP,' *Journal of Applied Econometrics*, 1995, 10, 109-125.

Tong, H., *Threshold Models in Non-Linear Time Series Analysis*, 1983, Springer-Verlag: New York.

Tsay, R.S., 'Testing and modelling Threshold Autoregressive Processes,' *Journal of the American Statistical Association*, 1989, 82, 590-604.

Yadav, P.K., P.F. Pope, and K. Paudyal, 'Threshold Autoregressive modelling in Finance: The Price Differences of Equivalent Assets,' *Mathematical Finance*, 1994, 4, 205-221.

Zellner, A. *An Introduction to Bayesian Inference in Economics*, 1971, Wiley, New York.

Appendix

The Likelihood Function, Prior and Posterior Distributions

The likelihood function for the threshold ECM is given by the product of R individual functions stemming from the usual likelihood function for a VAR model.

$$L\left(\left\{B^{(j)},\Sigma^{(j)}, j=1,\ldots,R\right\},r,d\right) \propto \prod_{j=1}^{3}(2\pi)^{-mN^{(j)}/2}\left|\Sigma^{(j)}\right|^{-N^{(j)}/2} \cdot \exp\left\{-\tfrac{1}{2}tr[S^{(j)}+(B^{(j)}-\hat{B}^{(j)})' X^{(j)\prime} X^{(j)}(B^{(j)}-\hat{B}^{(j)})]\Sigma^{(j)^{-1}}\right\} \tag{A1}$$

where $\hat{B}^{(j)}=\left(X^{(j)\prime}X^{(j)}\right)^{-1}X^{(j)\prime}Y^{(j)}$ is the OLS estimator of $B^{(j)}$, given r and d and

$$S^{(j)}=\left(Y^{(j)}-X^{(j)}\hat{B}^{(j)}\right)'\left(Y^{(j)}-X^{(j)}\hat{B}^{(j)}\right) \tag{A2}$$

is an m x m sample covariance matrix for regime $R^{(j)}$, given r and d. Here each element $b_i^{(j)} \in \mathcal{R}^1, \Sigma^{(j)}$ is an m x m positive definite matrix, $a_L \le r_1 < r_2 \le a_U$ and $d = 1, 2, \ldots, D$.

To complete a Bayesian analysis of our model, the specification of the joint prior distribution for our unknown parameters must be included. Our choice is of the form

$$p\left(\left\{B^{(j)},\Sigma^{(j)}, j=1,\ldots,R\right\},r,d\right) \propto \prod_{j=1}^{3}|\Sigma^{(j)}|^{-(m+1)/2}p(r,d). \tag{A3}$$

so that the joint prior for $B^{(j)}$ and $\Sigma^{(j)}$, conditional on r and d, is of the usual Jeffreys[6] form, assuming *a priori* independence of $B^{(j)}$ and $\Sigma^{(j)}$ (see Zellner, 1971). In addition, we let the marginal prior distribution for r and d be uniform over the appropriate region, so that

$$p(r,d) \propto 1 \text{ for } a_L = r_0 < r_1 < \ldots < r_R = a_U, \text{ and } d = 1,2,\ldots,D. \tag{A4}$$

Note that, under this specification, both the likelihood function, the prior density and, consequently, the posterior density are flat for any r^j in $[z_q, z_{q+1})$. Consequently, no distinction can be made with respect to threshold values between observed z values.

[6] The prior given is a Jeffreys prior for multiple regression as given in Zellner (1971). For discussion of a Jeffreys prior in the context of time series models, see Phillips (1991)

Under the above prior specification, the marginal posterior density for r and d can be obtained by integrating the product of the likelihood function in (A1) and the prior density given by (A3) and (A4), and integrating first with respect to $B^{(j)}$

$$p(\Sigma^{(j)},r,d|y) \propto \prod_{j=1}^{3}\int(2\pi)^{-mN^{(j)}/2}\left|\Sigma^{(j)}\right|^{-(N^{(j)}+m+1)/2} \cdot\exp\left\{-\tfrac{1}{2}tr\left\{\left(S^{(j)}+(B^{(j)}-\hat{B}^{(j)})'X^{(j)\prime}X^{(j)}(B^{(j)}-\hat{B}^{(j)})\right)\Sigma^{(j)^{-1}}\right\}\right\}dB^{(j)} \tag{A6}$$

so

$$p(\Sigma^{(j)},r,d|y) \propto \prod_{j=1}^{3}(2\pi)^{-mv^{(j)}/2}\left|\Sigma^{(j)}\right|^{-v/2}\left|X^{(j)\prime}X^{(j)}\right|^{mk^{(j)}/2}\exp\left\{-\tfrac{1}{2}trS^{(j)}\Sigma^{(j)^{-1}}\right\} \tag{A7}$$

and hence, for given r and d, the above is proportional to an $IW_m(v^{(j)},S)$ density. Integrating (A7) with respect to $\Sigma^{(j)}$, the marginal posterior density of r and d yields

$$p(r,d|y) \propto \prod_{j=1}^{3}\pi^{-mv^{(j)}/2}\prod_{j=1}^{m}\Gamma\left[\left(v^{(j)}+1-i\right)/2\right]\left|S^{(j)}\right|^{-v^{(j)}/2}\left|X^{(j)\prime}X^{(j)}\right|^{-m/2} \tag{A8}$$

where r and d are constrained to $a_L = r_0 < r_1 < ... < r_R = a_U$, and $d = 1,2,...,D$.

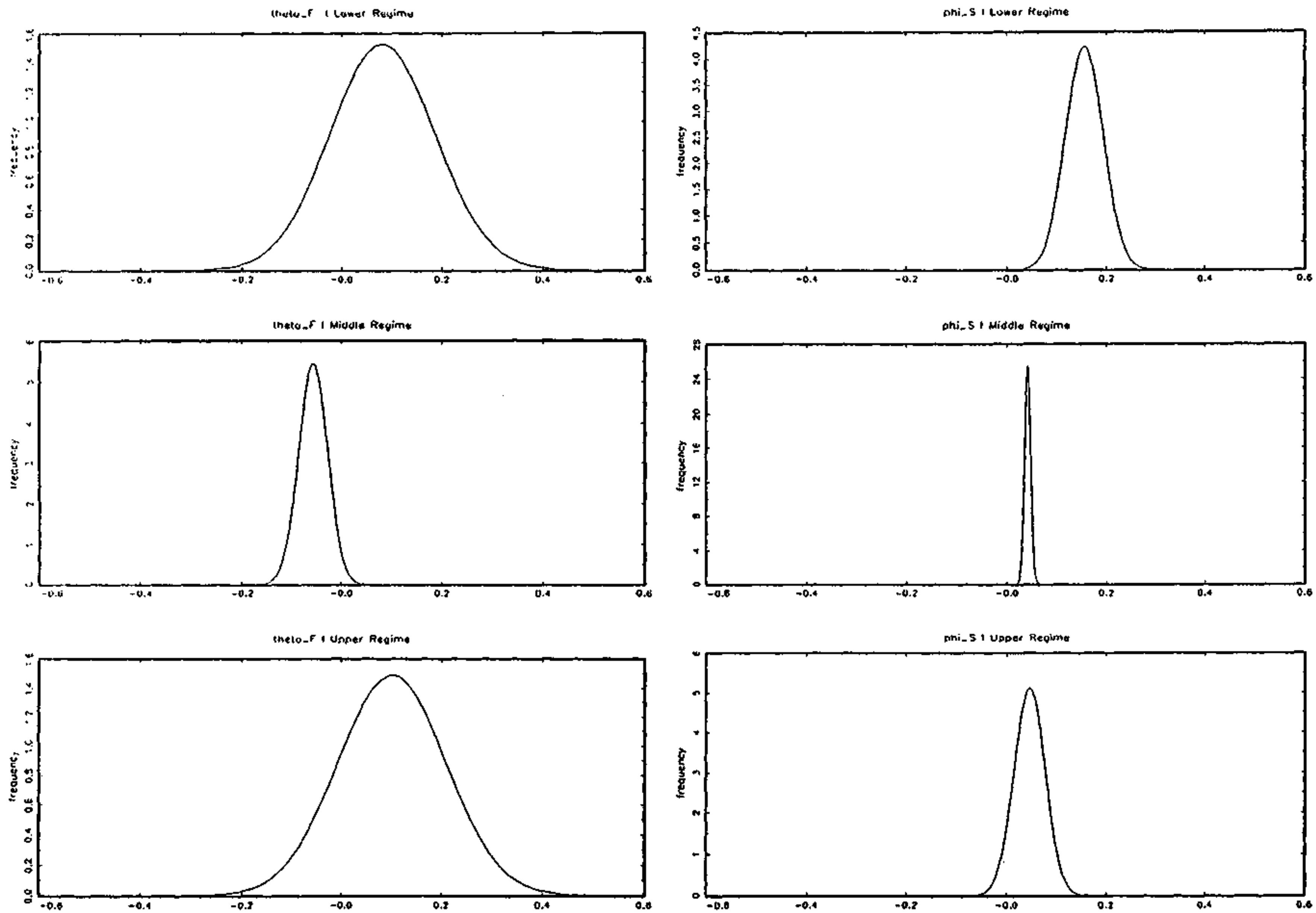

Figure 1. First-Order Cross-Autoregression Parameters

Figure 2. Joint Posterior Densities for the Gamma Parameters

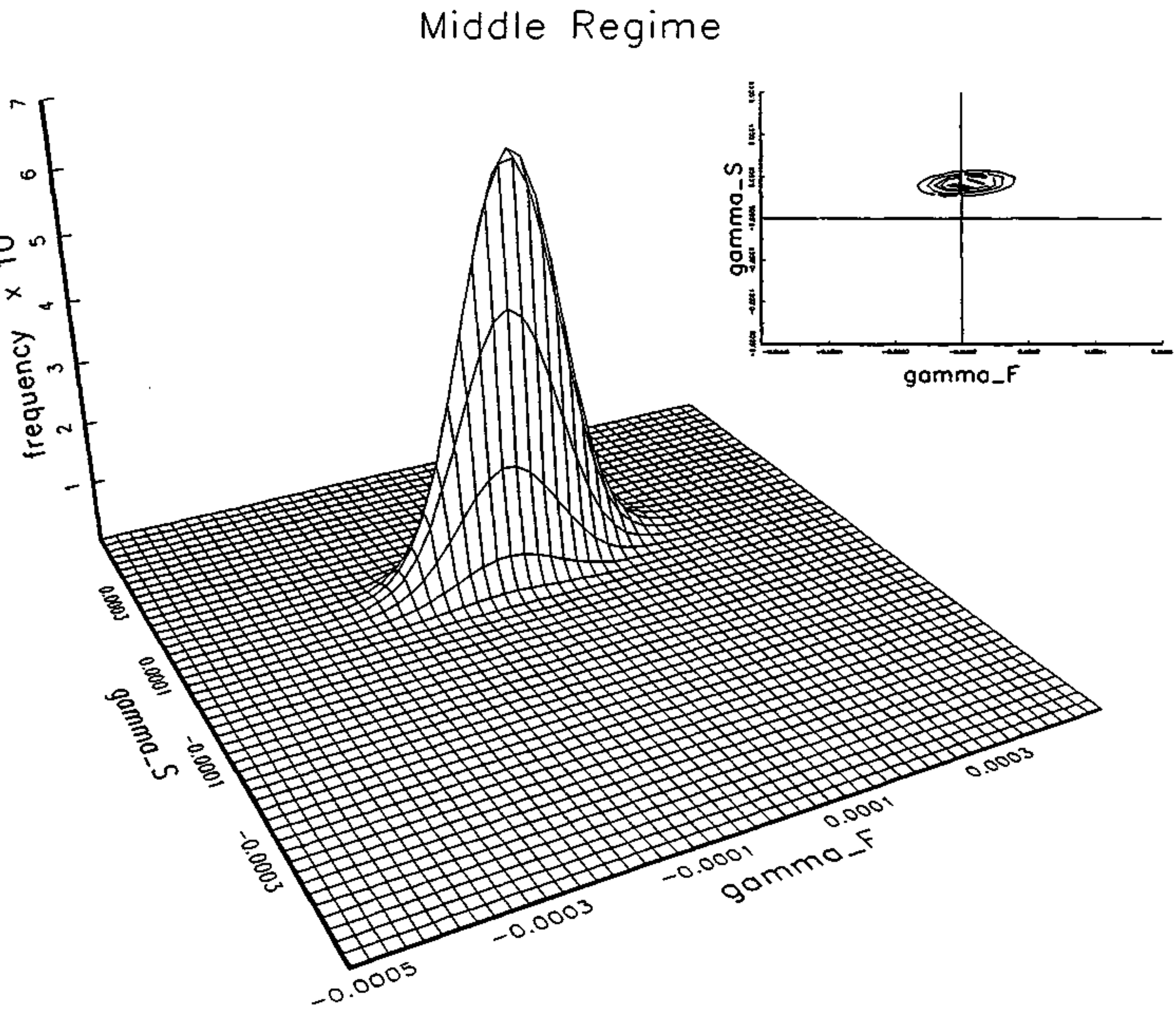

Figure 2. continued.

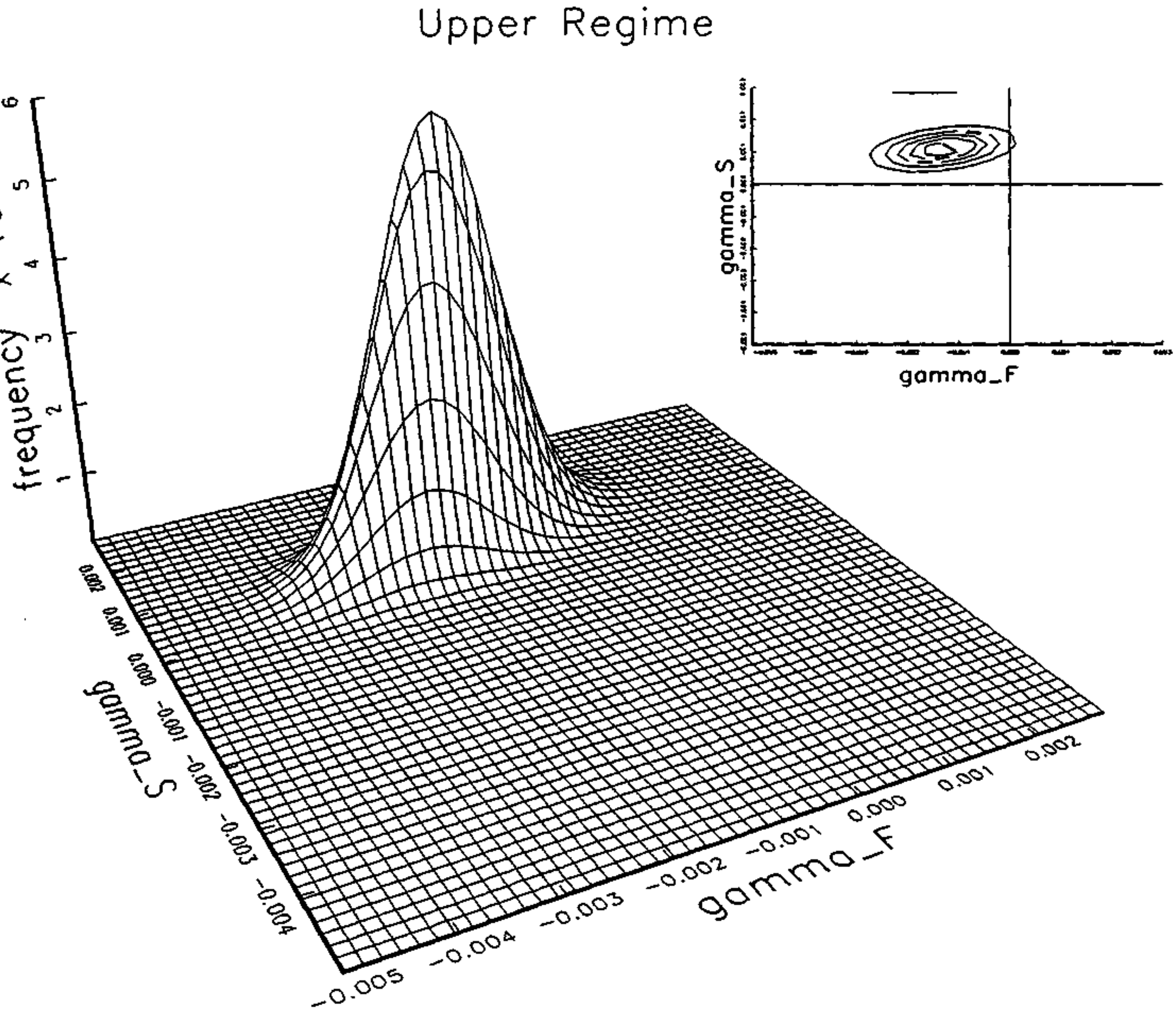

Figure 2. continued.