Financial Econometrics Research Centre

WORKING PAPERS SERIES

WP04-11

Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers

Lucio Sarno and Giorgio Valente

Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers*

Lucio Sarno
University of Warwick
and
Centre for Economic Policy Research (CEPR)

Giorgio Valente
University of Warwick

First version: April 2002 - Second revised version: October 2003

Abstract

This paper proposes a vector equilibrium correction model of stock returns that exploits the information in the futures market, while allowing for both regime-switching behavior and international spillovers across stock market indices. Using data for three major stock market indices since 1989, we find that: (i) in sample, our model outperforms several alternative models on the basis of standard statistical criteria; (ii) in out-of-sample forecasting, our model does not produce significant gains in terms of point forecasts relative to more parsimonious alternative specifications, but it does so both in terms of market timing ability and in density forecasting performance. The economic value of the density forecasts is illustrated with an application to a simple risk management exercise.

JEL classification: G10; G13.
Keywords: stock returns; futures; forecasting; nonlinearity; regime switching.

***Acknowledgments:* This paper was partly written while Lucio Sarno was a Visiting Scholar at the Federal Reserve Bank of St. Louis, the International Monetary Fund and the Central Bank of Norway, whose hospitality is gratefully acknowledged. We are grateful to John Geweke (editor) and three anonymous referees for thoughtful and constructive comments on previous drafts which led to a substantially improved paper. We are also thankful to Abhay Abhyankar, Gianni Amisano, Dick van Dijk, Oyvind Eitrheim, Philip Franses, Lutz Kilian, Hans-Martin Krolzig, Mark Taylor, Ken Wallis and participants to the 2002 Royal Economic Society Annual Conference at the University of Warwick for useful conversations or comments on previous drafts. Financial support from the Economic and Social Research Council is gratefully acknowledged. The authors alone are responsible for the views expressed and any errors that may remain.

1 Introduction

A large body of research on modeling and forecasting stock returns has investigated the relationship between spot and futures prices in stock index futures markets. In particular, a number of empirical studies have focused on the persistence of deviations from the cost of carry and have investigated the relationship between spot and futures prices in the context of vector autoregressions using cointegration or equilibrium correction models (see Dwyer, Locke and Yu, 1996; Neely and Weller, 2000, and the references therein). The rationale underlying this line of research is that the cost of carry model and variants of it predict that spot and futures prices cointegrate and their long-run relationship is characterized by a long-run equilibrium defined by the futures basis, implying both mean reversion in the basis and the existence of a vector equilibrium correction model (VECM) for spot and futures prices.[1] This literature, discussed in greater detail in the next section, has generally reported evidence that the futures market contains valuable information for modeling and/or forecasting stock returns.

A related line of research emphasizes that trading activity does not take place for one index per unit of time (e.g. see Eun and Shin, 1989; Engle and Susmel, 1994; Koutmos and Booth, 1995; Lee, 1995; Karoly and Stulz, 1996). Indeed, it is more likely that traders place orders and take positions simultaneously using different indices given that stock and futures markets for different indices are closely linked by both hedging activities and cross-market arbitrage. This may generate comovements across stock market indices and, in turn, the cross-correlation between different indices may be potentially very useful in improving empirical models of stock returns. In particular, it seems possible that, in the unknown dynamic model governing the relationship between futures and stock prices, stock returns for a particular index respond not only to the disequilibrium in the relevant stock index market but also to disequilibria in stock index markets that are linked to the relevant stock index by hedging activities and cross-market arbitrage (e.g. Ang and Bekaert, 2001; Goetzmann, Li and Rouwenhorst, 2001; Martens and Poon, 2001).[2]

Alongside the work on modeling and forecasting stock prices and returns, another strand of the literature has developed where increasingly strong evidence of nonlinearities in stock price movements has been documented. One element of this has been the mounting evidence that the conditional distribution of stock returns is well described by a mixture of normal distributions (e.g. see Rydén, Teräsvirta and Åsbrink, 1998, and the references therein) and that, consequently, a Markov switching model may be a logical characterization of stock returns behavior (e.g. see, *inter alia*, LeBaron, 1992; Hamilton and Susmel, 1994; Hamilton and Lin, 1996; Ramchand and Susmel, 1998a,b; Rydén, Teräsvirta and Åsbrink, 1998; Susmel, 1999; Perez-Quiros and Timmermann, 2001). Also, not only Markov-switching models fit stock returns data well, but they have sometimes been proved to produce superior forecasts to several alternative conventional models of stock returns (e.g. see Hamilton and Susmel, 1994; Hamilton and Lin, 1996).[3]

In this paper, we tie together these somewhat different, albeit related, strands of research.

[1] Several authors have recently begun to use the term 'equilibrium correction' instead of the traditional 'error correction' as the latter term now seems to have a different meaning in some recent theories of economic forecasting (e.g. see Clements and Hendry, 1998, p. 18). Since the term 'equilibrium correction' conveys the idea of the adjustment considered in the present context quite well, we use this term below.

[2] For example, Ang and Bekaert (2001) find that cross-country predictability is stronger than predictability using local instruments. Goetzmann, Li and Rouwenhorst (2001) document the correlation structure of several major equity returns over 150 years.

[3] Other studies in this literature have provided ample empirical evidence that the dynamic relationship linking stock and futures prices may display significant nonlinearities that can be well characterized using threshold models of various sort. These nonlinearities are rationalized on the basis of factors such as non-zero transactions

In particular, we investigate whether allowing for nonlinearities and international spillovers in the underlying data-generating process for a VECM that links spot and futures prices yields an improvement, in terms of both in-sample fit and out-of-sample forecasting, over models of stock returns that do not allow for nonlinearities and/or international spillovers. This is done through estimating a fairly general Markov-switching VECM (MS-VECM) for stock and futures prices that is based on an extension of Markovian regime shifts to a nonstationary framework. Given the evidence of significant regime-switching behavior in stock returns and the evidence on international cross-correlations of stock returns discussed above, this seems a natural way to extend current econometric procedures applied to stock returns modeling and forecasting, even though this involves estimating and forecasting from a sophisticated multivariate nonlinear model.

Using weekly data since 1989 for three major stock market indices - the S&P 500, the NIKKEI 225 and the FTSE 100 indices - we confirm that the futures market does contain some valuable information to explain stock returns in a linear VECM framework. However, we show that allowing for nonlinearities and for international spillovers in an MS-VECM results in a superior empirical model which explains a sizable proportion of the stock returns examined over our sample. We then compare the performance of our proposed model to several alternative linear and nonlinear models in an out-of-sample forecasting exercise. The evaluation of the relative performance is based on conventional statistical criteria for point forecasting performance, on tests of market timing, and on density forecasting evaluations. In fact, we argue and provide evidence that density forecast accuracy is more appropriate for evaluating our competing models since stock returns are non-normally distributed and we are considering nonlinear models consistent with non-normal densities (see, *inter alia*, Diebold, Gunther and Tay, 1998; Granger and Pesaran, 1999; Tay and Wallis, 2000; Timmermann, 2000).

To anticipate our forecasting results, we find that the MS-VECM that allows for international spillovers does not outperform the competing models examined in terms of point forecasting performance. However, our model significantly outperforms all of the competing models both in terms of market timing ability and in terms of density forecasting performance in that it generates predictive densities that are much closer to the true predictive density of the data.[4] Overall, these results suggest that, while the statistical performance of the linear and nonlinear models examined in this paper differs little in terms of conditional mean, investigation of hit rates and market timing ability suggests that the most general nonlinear model proposed performs better in forecasting the direction of future stock returns. In addition, inspection of the predictive densities implied by the various models also allows us to discriminate between models, indicating that both multiple regimes and the allowance for international spillovers are important ingredients for a model to produce satisfactory predictive densities. This implies that, although the various models examined do not differ statistically in terms of their predictive performance with respect to the conditional mean, the most general nonlinear model proposed provides a better characterization of the uncertainty surrounding the point forecasts.

We illustrate the practical importance of our results on density forecasting with a simple

---

costs or infrequent trading or simply the existence of regime shifts in the dynamic adjustment of stock and futures price changes towards their long-run equilibrium values (e.g. see, *inter alia*, Yadav, Pope and Paudyal, 1994; Dwyer, Locke and Yu, 1996; Martens, Kofman and Vorst, 1998; Gao and Wang, 1999; Aslanidis, Osborn and Sensier, 2002). See also Andreou, Osborn and Sensier (2000), Perez-Quiros and Timmermann (2001) and Timmermann (2001).

[4]By true predictive density of the data we mean the density of the data over the chosen forecast period. Therefore, no forecast is in fact carried out in this case, and the term 'predictive' simply refers to the fact that the density in question does not refer to the full sample but only to the forecast period. Also note that we use the terms 'predictive density' and 'forecast density' interchangeably below.

application to a risk management exercise. In recent years, trading accounts at large financial institutions have shown a dramatic growth and become increasingly more complex. Partly in response to this trend, major trading institutions have developed risk measurement models designed to manage risk. These models generally employ the Value-at-Risk (VaR) methodology, where VaR is defined as the expected maximum loss over a target horizon within a given confidence interval (Jorion, 2000; Basak and Shapiro, 2001).[5] In our simple application we analyze the out-of-sample forecasting performance of our proposed empirical models of stock returns, investigating the implications of these forecasts for a risk manager who has to quantify the risk associated with holding the stock indices in question over a one-week horizon. This application further illustrates how the MS-VECM that allows for international spillovers captures satisfactorily the higher moments of the predictive distribution of stock returns, generating VaRs that estimate the probability of large losses better than other competing models. Put another way, our findings indicate that better density forecasts of stock returns, of the type recorded by the most general regime-switching model considered in this paper, can potentially lead to substantial improvements in risk management and, more precisely, to better estimates of downside risk.

The remainder of the paper is set out as follows. In Section 2 we describe our empirical framework for modeling stock and futures prices allowing for international spillovers and nonlinear dynamics. We also briefly set out the econometrics of Markov-switching multivariate models as applied to nonstationary processes and cointegrated systems. In Section 3 we describe the data and report the results relating to the in-sample empirical analysis. In Section 4 we report our forecasting results, including evidence on point forecast accuracy, market timing ability, density forecast accuracy and an illustrative application to risk management aimed at investigating the importance of density forecasting in the context of stock returns. A final section concludes.

2 Modeling stock returns: an empirical framework

In this section we outline our empirical framework for modeling stock returns, which we apply to our data in the subsequent sections. First, we use a conventional cost of carry model to show that futures and stock prices must be cointegrated and, therefore, linked by a VECM that can be used both to explain and forecast stock returns. Second, we generalize the VECM linking stock and futures prices to take into account potentially important regime switches of the kind reported by a large empirical literature. Third, we further generalize our empirical framework by also taking into account the observed cross-correlations between major stock market indices, which leads us to consider a panel of VECMs which explicitly allows for both regime shifts and international spillovers across major stock market indices.

2.1 The information in the futures market

A useful starting point for building an empirical framework to model stock returns is the relationship between stock prices and stock futures prices, as described by a conventional cost of carry model with no transaction costs:

[5] More formally, VaR is an interval forecast, typically a one-sided 95 or 99 percent interval of the distribution of expected wealth or returns.

$$F(t,T) = S(t)\exp\left\{\sum_{k=1}^{T-t} c(t+k)\right\}, \tag{1}$$

where $S(t)$ is the stock index price, $F(t,T)$ is the futures price at time t for delivery of the stock at time $T \geq t$ and $c(t+k)$ denotes the expected net cost of carry for period $t+k$. Taking logs, equation (1) can be rewritten as

$$\log F(t,T) - \log S(t) = \sum_{k=1}^{T-t} c(t+k), \tag{2}$$

where $\log F(t,T) - \log S(t)$ is the log-basis. Following Low, Muthuswamy, Sakar and Terry (hereafter LMST, 2002), suppose that market expectations about the cost of carry for each period are drawn from independent and identical normal distributions, each with mean $\overline{c}$ and variance σ_c^2. Then the log-basis will be normally distributed with mean $\overline{c}(T-t)$ and variance $\sigma_c^2(T-t)$. This implies that both the first and second moments of the log-basis will be functions of the time to maturity $(T-t)$ (see LMST, 2002).[6] If the expected cost of carry for each period has a stationary distribution, then equation (2) implies cointegration between futures and spot prices with the cointegrating relationship given by

$$z_t = \log F(t,T) - \log S(t) - \overline{c}(T-t). \tag{3}$$

Equation (3) implies that the futures and the underlying spot prices cointegrate with a cointegrating vector which differs from the usual cointegrating vectors investigated in the empirical literature on the cost of carry model (e.g. Lien and Lou, 1993; 1994; Kroner and Sultan, 1993, Gagnon and Lypny, 1995) as a result of the presence of the term $\overline{c}(T-t)$. Given equation (3), z_t may be seen as the stationary deviation from the cost-of-carry model. In turn, the Granger Representation Theorem (Engle and Granger, 1987) implies that the futures and stock prices must possess a VECM representation where the log-basis adjusted for the time to maturity term (z_t) plays the part of the equilibrium error.[7] We exploit this framework and use exactly a VECM representation to demonstrate that valuable information may be extracted from the futures market in order to explain and forecast stock returns (LMST, 2002).

[6] Strictly speaking, the cost of carry model applies to forwards, not futures. In the case of futures, c is explained by time-varying interest rates and dividend yields. Given data at weekly frequency on dividend yields, one could calculate the adjusted log-basis using interest rates and dividend yields to match the remaining time to maturity. However, weekly data on dividend yields are typically difficult to obtain and need to be interpolated (under an assumed process for dividends), potentially reducing the accuracy of the basis calculations. While some studies use the 'de-meaning' method per day (not using interest rates and dividends, and relying instead on a large number of intraday observations for each day) and assume that the time to maturity is approximately constant (e.g. Dwyer, Locke and Yu, 1996), this approach cannot be applied to an entire data set of weekly data. Given these difficulties, we follow the method of LMST (2002) in our calculation of the adjusted log-basis, where we correct for the time to maturity $T-t$. See also Section 3.1.

[7] The precise definition of cointegration requires the cointegrating vector to be covariance stationary. Because equation (3) implies that the variance of the cointegrating vector will be a function of the time to maturity, the futures and underlying spot price cannot be cointegrated in a strict sense. However, Hansen (1992a) shows that much of the statistical theory developed under the strict definition of cointegration still holds when heteroskedasticity is permitted in the cointegrating vector. See LMST (2002) for a detailed discussion of the cointegrating properties of the cost of carry model in this context.

2.2 Regime-switching equilibrium correction in stock index futures markets

A large literature has documented evidence of nonlinearities in stock returns. One element of this has been the mounting evidence that the conditional distribution of stock returns is well described by a mixture of normal distributions (e.g. see Rydén, Teräsvirta and Åsbrink, 1998, and the references therein) and that, consequently, a Markov switching model may be a logical characterization of stock returns behavior (e.g. see, *inter alia*, LeBaron, 1992; Hamilton and Susmel, 1994; Hamilton and Lin, 1996; Ramchand and Susmel, 1998a,b; Rydén, Teräsvirta and Å sbrink, 1998; Susmel, 1999). In fact, the relevant literature suggests that not only Markov-switching models fit stock returns data well, but they may perform satisfactorily in forecasting (e.g. see Hamilton and Susmel, 1994; Hamilton and Lin, 1996).

In the present paper, we investigate whether allowing for regime-switching in the VECM implied by the framework described in the previous subsection yields a superior model of stock returns relative to several alternative specifications. This is done through estimating a fairly general MS-VECM for stock and futures prices which is based on an extension of Markovian regime shifts to a nonstationary framework. In the rest of this subsection we outline the econometric procedure employed in order to model regime shifts in the dynamic relationship between stock and futures prices. The procedure essentially extends Hamilton's (1988, 1989) Markov-switching regime framework to nonstationary systems, allowing us to apply it to cointegrated vector autoregressive (VAR) and VECM systems (see Krolzig, 1997, 1999, 2000).

Consider the following M-regime p-th order Markov-switching vector autoregression (MS(M)-VAR(p)) which allows for regime shifts in the intercept term:

$$y_t = \nu(\omega_t) + \sum_{i=1}^{p} \Pi_i y_{t-i} + \varepsilon_t, \tag{4}$$

where y_t is a K-dimensional observed time series vector, $y_t = [y_{1t}, y_{2t}, \ldots, y_{Kt}]'$; $\nu(\omega_t) = [\nu_1(\omega_t), \nu_2(\omega_t), \ldots, \nu_K(\omega_t)]'$ is a K-dimensional column vector of regime-dependent intercept terms; the Π_i's are $K \times K$ matrices of parameters; $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Kt}]'$ is a K-dimensional vector of Gaussian white noise processes with covariance matrix Σ, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma)$. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $\omega_t \in \{1, \ldots, M\}$, governed by the transition probabilities $p_{ij} = \Pr(\omega_{t+1} = j \mid \omega_t = i)$, and $\sum_{j=1}^{M} p_{ij} = 1 \; \forall i, j \in \{1, \ldots, M\}$.

A standard case in economics and finance is that y_t is nonstationary but first-difference stationary, i.e. $y_t \sim I(1)$. Then, given $y_t \sim I(1)$, there may be up to $K-1$ linearly independent cointegrating relationships, which represent the long-run equilibrium of the system, and the equilibrium error (the deviation from long-run equilibrium) is measured by the stationary stochastic process $u_t = \alpha' y_t - \beta$ (Granger, 1986; Engle and Granger, 1987). If indeed there is cointegration, the cointegrated MS-VAR (4) implies an MS-VECM of the form:

$$\Delta y_t = \nu(\omega_t) + \sum_{i=1}^{p-1} \Lambda_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t, \tag{5}$$

where $\Lambda_i = -\sum_{j=i+1}^{p} \Pi_j$ are matrices of parameters, and $\Pi = \sum_{i=1}^{p} \Pi_i - \mathbf{I}$ is the long-run impact matrix whose rank r determines the number of cointegrating vectors (e.g. Johansen, 1988, 1991).[8]

[8] In this section it is assumed that $0 < r < K$, implying that y_t is neither purely difference-stationary (i.e. $r = 0$) nor a stationary vector (i.e. $r = K$).

Although, for expositional purposes, we have outlined the MS-VECM framework for the case of regime shifts in the intercept alone, shifts may be allowed for elsewhere. The present application focuses on a multivariate model comprising, for each of the three major stock index markets analyzed, the futures price and the stock price (hence $y_t = [f_t, s_t]'$) where f_t and s_t denote the logarithmic futures and stock prices respectively. As discussed in Section 3 below, in our empirical work, after considerable experimentation, we selected a specification of the MS-VECM which allows for regime shifts in the intercept, the autoregressive structure and the variance-covariance matrix. This model, the Markov-Switching-Intercept-Autoregressive-Heteroskedastic-VECM or MSIAH-VECM, may be written as follows:

$$\Delta y_t = \nu(\omega_t) + \sum_{i=1}^{p-1} \Lambda_i(\omega_t)\Delta y_{t-i} + \alpha(\omega_t) z_{t-1} + u_t, \tag{6}$$

where $\alpha(\omega_t)$ is a regime-switching vector of speed of adjustment parameters and is part of the long-run impact matrix $\Pi(\omega_t) = \alpha(\omega_t)\beta'$, where β' is the cointegrating vector which generates a cointegrating relationship of the form given in equation (3); the equilibrium correction term z_{t-1} is as defined in equation (3); $u \sim NIID\,[\mathbf{0}, \Sigma(\omega_t)]$; and $\omega_t \in \{1, \ldots, M\}$. Intuitively, the shifts in the variance-covariance matrix allow us to capture the well-documented heteroskedasticity of stock returns over the sample examined. On the other hand, the need for shifts in the intercept and the autoregressive structure is consistent with the well-known evidence that analyses of forecasting that implicitly rule out structural breaks and regime shifts in the parameters ignore an aspect that may be responsible for a large number of episodes of predictive failure (e.g. Clements and Hendry, 1996). These corrections therefore offer greater protection against unforeseen regime shifts, potentially enhancing the forecasting performance of the model.

An MS-VECM can be estimated using a two-stage maximum likelihood procedure. The first stage essentially consists of the implementation of the Johansen (1988, 1991) maximum likelihood cointegration procedure in order to test for the number of cointegrating relationships in the system and to estimate the cointegration matrix. In fact, in the first stage use of the conventional Johansen procedure is valid without modeling the Markovian regime shifts explicitly (see Saikkonen, 1992; Saikkonen and Luukkonen, 1997). The second stage then consists of the implementation of an expectation-maximization (EM) algorithm for maximum likelihood estimation which yields estimates of the remaining parameters of the model (Dempster, Laird and Rubin, 1977; Hamilton, 1990; Kim and Nelson, 1999; Krolzig, 1999).

2.3 Separation and cointegration in modeling stock returns

Although conventional time series models employed to explain or forecast stock returns treat a particular asset or index in isolation, a vast literature in finance has pointed out that trading activity does not take place for one index per unit of time (see, *inter alia,* Eun and Shin, 1989; Engle and Susmel, 1994; Koutmos and Booth, 1995; Lee, 1995; Karoly and Stulz, 1996). This literature generally emphasizes that hedging activities and cross-market arbitrage may generate comovements across different stock market indices (Martens and Poon, 2001; Ang and Bekaert, 2001; Goetzmann, Li and Rouwenhorst, 2001) and, in turn, the correlation between different indices may be potentially very useful in improving empirical models of stock returns. In particular, it is possible that, in a VECM for futures and stock prices, stock price changes respond not only to the disequilibrium in the relevant stock index market but also to disequilibria in stock index markets that are linked to the relevant stock index.

This line of reasoning suggests the possibility of enriching our MS-VECM framework by allowing for spillovers through the equilibrium correction terms, that is the possibility that equilibrium correction terms from one cointegrating relationship for a particular stock market index may have explanatory power in the equilibrium correction equation driving the returns of another stock market index. This approach is consistent with the notion of separation and cointegration - popularized by Konishi and Granger (1993), Konishi (1993), Granger and Swanson (1996) and Granger and Haldrup (1997) - which therefore provides a useful way of describing formally the above ideas.

Consider, for example, the MS-VECM (5) and define an n-dimensional cointegrated vector $Y_t = \left[y_t^1, y_t^2, y_t^3\right]'$, where $y_t^j = [f_t^j, s_t^j]'$ for $j = 1, 2, 3$ is of dimension of n_j (i.e. $n = n_1 + n_2 + n_3$) and y_t^1, y_t^2 and y_t^3 have no variable in common. We can then generalize equation (5) to a VECM that exploits the information in the futures market while also allowing for both regime shifts and international spillovers. This VECM may be written as follows:[9]

$$\Delta Y_t = \boldsymbol{\nu}(\omega_t) + \sum_{i=1}^{p-1} \boldsymbol{\Lambda}_i \Delta Y_{t-i} + \boldsymbol{\alpha}\boldsymbol{\beta}' Y_{t-1} + \varepsilon_t, \tag{7}$$

where $\boldsymbol{\Lambda}_i$ is an $n \times n$ matrix of autoregressive parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ denote the $n \times r$ loading matrix and the $r \times n$ cointegration matrix (or matrix of cointegrating vectors) respectively, and r is the cointegration rank. The cointegration matrix $\boldsymbol{\beta}'$ can be factorized as

$$\boldsymbol{\beta}' = \begin{bmatrix} \beta'_{11} & 0 & 0 \\ 0 & \beta'_{22} & 0 \\ 0 & 0 & \beta'_{33} \end{bmatrix} \tag{8}$$

where β'_{jj} is $r^j \times n_j$, for $j = 1, 2, 3$. The system is said to have *separate cointegration* with cointegration ranks for each subsystem given by n_1, n_2 and n_3 respectively. If we then factorize the loading matrix as follows

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix}, \tag{9}$$

where α_{jj} is $n_j \times r^j$ for $j = 1, 2, 3$, we have *type B-separation* or separation in the equilibrium correction. Finally, if we factorize the matrix $\boldsymbol{\Lambda}_i$ as

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \Lambda_{11}^j & 0 & 0 \\ 0 & \Lambda_{22}^j & 0 \\ 0 & 0 & \Lambda_{33}^j \end{bmatrix} \tag{10}$$

we have *type A-separation* or separation in the dynamic adjustment towards the long-run equilibrium defined for each y_t^j for $j = 1, 2, 3$ (e.g. Granger and Haldrup, 1997). If all of the conditions (8)-(10) hold there is *complete separation*, while if condition (8) is associated with either (9) or (10) we have *partial separation*.

Our earlier discussion on spillovers in the dynamics of stock returns is consistent with a situation where, although two or more different stock indices are 'separated in the long-run' (i.e. condition (8) holds), there may be important short-run relationships between them

[9] For ease of exposition, in this subsection we ignore the fact that cointegration between futures and spot prices is consistent with a relationship of the form (3), namely $\log F(t, T) - \log S(t) - \overline{c}(T - t)$, including a term which varies with the term to maturity $\overline{c}(T - t)$. In our empirical work below, however, we explicitly consider the latter term.

and, therefore, the deviation from the equilibrium relationship from one index may enter the equilibrium correction equation of another index (i.e. condition (9) does *not* hold).

This 'amalgamation' is applied to the case of cointegration analysis across different stock indices in the world economy, which seems intuitively appealing given the high degree of integration of global capital markets during the last fifteen years or so. In particular, our framework is consistent with a situation where, for any stock index k, a long-run equilibrium relationship is established in a static cointegrating equation involving stock and futures prices for index k, as predicted by the cost of carry model. Hence, stock and futures prices for any other index $j \neq k$ do not enter the long-run cointegrating equation defining the equilibrium value of the stock price of index k. Despite long-run separation (that is the equilibrium value of the stock price of any index k is *fully* determined by the equilibrium relationship between stock and futures prices of the index k itself), however, the individual short-run relationships may be characterized by the equilibrium error from one equation entering another equilibrium correction equation of the system. This is the approach followed below, when we estimate a nonlinear MS-VECM where, for each stock index examined, the lagged deviation from equilibrium (equilibrium correction term) in other stock indices is allowed to enter the equilibrium correction equation in addition to the own-index lagged deviation from equilibrium (equilibrium correction term) in order to exploit the information content of international spillovers.

3 Empirical analysis I: modeling[10]

3.1 Data and preliminary statistics

The data set comprises weekly time series on prices of futures contracts written on the S&P 500, the NIKKEI 225 and the FTSE 100 indices, as well as price levels of the corresponding underlying cash indices. The data set is obtained from *Datastream*. Specifically, we use price levels of each stock index and corresponding futures contracts at the close of trade of every day. The data is collected to coincide with the length of the available futures contract. The futures data are constructed according to standard conventions (e.g. Ahn, Boudoukh, Richardson and Whitelaw, 2002). In particular, a single time series of future prices is spliced together from individual futures contract prices. For liquidity, the nearest contract's prices are used until the first day of the expiration month, then the next nearest is used. The adjusted log-basis has been constructed as in LMST (2002). We used equation (3) to calculate the log-basis adjusted for the time-to-maturity of each futures contract. In practice, for each stock index, we regressed $f_t - s_t$ on the time to maturity $(T - t)$ of each futures contract. We shall use the residual as the equilibrium correction term in our VECM estimation. All of the series considered have initially been constructed from daily data. We then obtained the weekly series from the daily series by using Wednesday prices, or Thursday prices when Wednesday prices were unavailable, in order to avoid potential weekend price effects (French, 1980; Gibbons and Hess, 1981; LMST, 2002).

The sample period examined spans from January 1989 to December 2002. We choose this sample period for two reasons. First, the NIKKEI 225 stock index futures was first traded on September 1988 in the Osaka Stock Exchange (OSE).[11] Second, given the focus of the

[10] Our data and programs are available upon request.

[11] More precisely, NIKKEI 225 futures contracts were first traded in 1986 in the Singapore International Monetary Exchange (SIMEX). Since NIKKEI 225 futures contracts are more actively traded in the OSE than the SIMEX we prefer to use the OSE data (see Pan and Hsueh, 1998, for further discussion of the institutional details of trading the NIKKEI 225 stock index futures contracts).

present paper on investigating the importance of allowing for nonlinearity (regime switching) in modeling stock returns, using data after the 1987 crash should reduce the risk that the nonlinearity detected and modeled in the empirical analysis could be determined by or attributed to a unique and perhaps exceptional event occurred over the sample. In our empirical work, we carried out estimations over the period January 1989-December 1998, reserving the last four years of data for out-of-sample forecasting tests.

A number of related studies motivated by microstructure considerations or focusing on modeling intraday or short-lived arbitrage have used intraday data at various intervals or daily data - e.g. Miller, Muthuswamy and Whaley (1994) and Dwyer, Locke and Yu (1996) use 15- and 5-minute intervals respectively. In order to reduce the noise element in the data, we choose to employ data at weekly frequency. However, we carried out a fraction of the estimation work reported below also using daily data. These estimation results were qualitatively identical, suggesting that aggregation from daily to weekly may not have particularly important effects on the regime-switching properties of our stock returns data.[12]

It is worth noting that in estimating and forecasting from our (linear or nonlinear) VECMs for stock and futures prices (recorded at the close of trade for each Wednesday), it is possible to estimate any of the VECMs described in Section 2 having the relevant information at time t in order to forecast stock returns for each stock index considered at time $t+1$. Also, note that the subscript t always refers to close of trading in the S&P 500 (i.e. t=15.15 Chicago time). This is important for the following reasons. The trading hours of the three markets examined, in local times, are as follows: 9.00-15.10 for the NIKKEI 225, which is the first market to open among the three considered; 8.00-17.30 for the FTSE 100, which is the second market to open; and 8.30-15.15 for the S&P 500, which the last market to open. The relevant time with respect to Greenwich Mean Time (GMT) is +9 for the NIKKEI 225; 0 for the FTSE 100; and -6 for the S&P 500.[13] For our VECMs to be used as forecasting models, one must have, at time t, defined as above, information available on each of the three stock returns (and lagged values) as well as information on each of the three futures bases. This is indeed the case since, standing at time t (i.e. close of S&P 500), the local time in Osaka is 6.15, that is 2.45 hours *before* the NIKKEI 225 opens again for trading. Hence, at time t, as defined above, one can estimate any of the VECMs discussed in Section 2, using the publicly available information on stock prices and returns and the futures basis at time t for each of the NIKKEI 225, FTSE 100, and S&P 500, and it is possible to produce one-step-ahead (one-week-ahead) forecasts of the three stock returns examined at time $t+1$.[14] The bottom line of the above discussion is that at the close of the S&P 500 (time t), one has an information set comprising $f_t - s_t$, Δf_t and Δs_t (as well as their lags) for each of the three indices examined; this is the information set needed to employ the VECMs discussed in Section 2 as forecasting models of stock returns of the NIKKEI 225, FTSE 100 and S&P 500 at time $t+1$.

Table 1 provides summary statistics of the logarithm of the futures price, f_t and the logarithm of the spot price, s_t. As one would expect, for each stock index, the first moment of the

[12] Nevertheless, given the high computational burden of executing the work discussed below in the forecasting exercise, using weekly (rather than daily) data allowed us to be more ambitious in terms of the amount of overall empirical work carried out.

[13] This is the time with respect to GMT ignoring Daylight Saving Time (DST). Taking into account DST only changes our calculations by one hour at some point during the summer when both in the US and UK markets the time will be as given above plus one hour (essentially giving the forecaster one more hour available to run his/her model).

[14] Also, we were careful in avoiding the problems caused by nonsynchronous market closure. Specifically, given that the futures market and spot market cease trading at slightly different times, we use spot and futures data recorded at the close of the market which closes first.

futures price is larger than the first moment of the spot price (although it is not the case that $f_t > s_t$ at each point in time), while the second moment of the spot price is larger than the second moment of the futures price, suggesting that the futures price is larger on average and less volatile than the spot price. The partial autocorrelation functions, reported in Table 1 up to order 12, suggest that each spot and futures price examined displays very strong first-order serial correlation, while none of these series appears to be significantly serially correlated at higher lags. This is confirmed by the visual evidence provided in Figure 1, which plots the time series to be predicted, namely Δs_t, over the full sample period.[15]

3.2 Cointegration tests and linear dynamic modeling[16]

We tested for cointegration between f_t and s_t by employing the Johansen procedure in a VAR which allows for an unrestricted constant term. Both Johansen likelihood ratio (LR) test statistics (based on the maximum eigenvalue and on the trace of the stochastic matrix respectively) clearly suggested that a cointegrating relationship existed. Also, the hypothesis that the cointegrating parameter associated with s_t equals unity could not be rejected at conventional nominal levels of significance for each of the estimated VARs.[17] However, although these cointegration results prove that futures and spot prices cointegrate with a unity parameter, they do not provide us with the most appropriate equilibrium correction terms for estimating a VECM for Δf_t and Δs_t since the cointegrating relationship tested for does not allow for the time-varying nature of the cost of carry and the time-to-maturity effect discussed in Section 2.1. The equilibrium correction term we use in our VECM estimation is the adjusted log-basis calculated following LMST (2002), as discussed in Sections 2.1 and 3.1.

As a further preliminary to considering an MS-VECM, we estimated a standard linear bivariate VECM for Δf_t and Δs_t. Thus, using full-information maximum likelihood (FIML) methods, we estimated for each stock index a bivariate VECM of the form

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Lambda_i \Delta y_{t-i} + \alpha z_{t-1} + \varepsilon_t, \tag{11}$$

where $y_t = [f_t, s_t]'$, α is a 2×1 vector of speed of adjustment parameters, and z_{t-1} is as defined in equation (3). We allowed for a maximum lag length of five, which was the longest lag length selected by the Akaike information criterion (AIC) and the Bayes information criterion (BIC); in case of conflicting results from the AIC and BIC we chose the longest lag length. Employing the conventional general-to-specific procedure, we then obtained, for each stock index examined, fairly parsimonious models for Δf_t and Δs_t which display no residual serial correlation.

Further, in order to test for cointegration and separation of the type discussed in Section 2.3, we estimated the following model:

[15] As a preliminary exercise, we tested for unit root behavior of the (log) futures price and spot price time series by calculating standard augmented Dickey-Fuller test statistics. In keeping with a large number of studies and conventional finance theory, we were in each case unable to reject the unit root null hypothesis at conventional nominal levels of significance. On the other hand, differencing the series did appear to induce stationarity in each case. Overall, the unit root tests clearly indicate that both f_t and s_t are realizations from stochastic processes integrated of order one, which suggests that testing for cointegration between f_t and s_t is the logical next step.

[16] The empirical results discussed in this subsection, which are preliminary to the work carried out subsequently in the paper, are not reported to save space, but they are available from the authors upon request.

[17] LR tests of the hypothesis that the coefficient associated with s_t equals unity could not be rejected with p-values equal to 0.610, 0.572 and 0.607 for the S&P 500, the NIKKEI 225 and the FTSE 100 respectively.

$$\Delta Y_t = \boldsymbol{\nu} + \sum_{i=1}^{p-1} \boldsymbol{\Lambda}_i \Delta Y_{t-i} + \alpha Z_{t-1} + \boldsymbol{\varepsilon}_t, \tag{12}$$

where $Y_t = [f_t^{SP500}, s_t^{SP500}, f_t^{NK225}, s_t^{NK225}, f_t^{FTSE100}, s_t^{FTSE100}]'$, α is a 6×3 loading matrix, and $Z_{t-1} = [z_t^{SP500}, z_t^{NK225}, z_t^{FTSE100}]'$. We tested for *type B-separation* (separation in the equilibrium correction) by estimating model (12) and testing the zero restrictions in (9) using a standard likelihood ratio (LR) test. The results allow us to reject the zero restrictions under the null hypothesis (9) with a p-value of 10^{-16}, implying that there is no separation in the equilibrium correction, or put differently, that the disequilibrium (deviation of the basis from its equilibrium level) in one index influences the dynamics of stock returns of other indices.

As a check of adequateness of the models as well as an additional motivation for the need of employing a nonlinear model to characterize the dynamic relationship between stock and futures prices, however, we employed two fairly general tests for linearity of the residuals from the VECMs (11) and (12), namely Ramsey's (1969) RESET test and the Brock, Dechert and Sheinkman (BDS, 1991) test for the null hypothesis that the residuals from (11) and (12) are independent and identically distributed (iid) against an unspecified alternative. Application of both of these tests provided strong evidence that the linear VECMs (11) and (12) fail to capture important nonlinearities in the data generating process, as linearity is generally rejected with marginal significance levels (p-values) of virtually zero.[18]

3.3 MS-VECM estimation results

Next, we applied a "bottom-up" procedure designed to detect Markovian shifts in order to select the most adequate characterization of an M-regime p-th order MS-VECM for Δy_t of the form discussed in Section 2. However, for each MS-VECM estimated the assumption that the regime shifts affect only the intercept term of the VECM was found to be inappropriate. In fact, we checked in turn the relevance of regime-conditional heteroskedasticity and regime-conditional autoregressive structure. We then tested the hypothesis of no regime dependence using an LR test of the type suggested by Krolzig (1997, p. 136). The results suggest very strong rejections of the null of no regime dependence, clearly indicating that an MS-VECM that allows for shifts in the intercept, the autoregressive structure, the cointegrating matrix and the variance-covariance matrix, that is an MSIAH(M)-VECM(p), is the most appropriate model within its class in the present application. We also tested for the significance of the autoregressive structure and found that $p = 1$ is the lag length which better characterizes the dynamics of the series. For simplicity, we assume, as done in much recent literature on Markov-switching models (see, *inter alia*, Cecchetti, Pok-Sang and Mark, 1990, 2000; Hamilton and Lin, 1996; Ramchand and Susmel, 1998a,b; Perez-Quiros and Timmermann, 2001), the presence of two regimes for each stock index.

Thus, we selected and estimated a bivariate MSIAH(2)-VECM(1) for Δy_t of the form

$$\begin{aligned} \Delta y_t &= \nu(\omega_t) + \sum_{i=1}^{p-1} \Lambda_i(\omega_t) \Delta y_{t-i} + \alpha(\omega_t) z_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim NID[\mathbf{0}, \Sigma(\omega_t)] \qquad \omega_t = 1, 2 \end{aligned} \tag{13}$$

[18] However, we also used the linear VECMs to forecast the future stock price and compared these forecasts to the forecasts obtained from an MSIAH-VECM, as discussed below.

using the EM algorithm for maximum likelihood estimation discussed in Section 2. In order to test for cointegration and type B-separation we also estimated the following model

$$\begin{aligned}\Delta Y_t &= \boldsymbol{\nu}(\omega_t) + \sum_{i=1}^{p-1} \boldsymbol{\Lambda}_i(\omega_t)\,\Delta Y_{t-i} + \boldsymbol{\alpha}(\omega_t)\,Z_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim NID\left[\mathbf{0}, \boldsymbol{\Sigma}(\omega_t)\right] \qquad \omega_t = 1, \ldots, 2^3. \end{aligned} \tag{14}$$

Making no assumption on the relationship between the regime shifts implies that the number of regimes incorporated in model (14), and consequently the dimension of the transition matrix, is $2^3 = 8$ (see Hamilton and Lin, 1996; Krolzig, 1997) - for technical details see Appendix A.

We compute an LR test statistic for linearity (LR1), which essentially tests the hypothesis that the true model is a linear VECM against the alternative of the MSIAH-VECM, reported in Table 2. The test was first carried out using a lag length of five in each of the linear VECM and the MSIAH-VECM. Even by invoking the upper bound of Davies (1977, 1987), the linearity hypothesis is rejected very strongly, with a p-value of virtually zero, providing convincing evidence of the need of employing a regime-switching model.[19] Then, in order to test the best linear model for each stock index with the preferred lag length (which may be greater than unity) against the preferred MSIAH-VECM (which is always found to have only one lag), we also calculate J-tests of the type developed by Davidson and MacKinnon (1981) for the null hypothesis of equality of two non-nested models. The results, also reported in Table 2, confirm the rejection of the best fitting linear model against the MSIAH-VECM with one lag in each case. Moreover, even in the context of Markov-switching models, type B-separation is rejected by the data. In fact the likelihood ratio test (LR2) reported in the second column of Table 2 strongly rejects the null of separation in the equilibrium correction terms.

We also compute coefficients of determination $\overline{R}^2$, which were adjusted both for the bias towards preferring a larger model relative to a smaller one as well as for the fact that the model allows for regime dependence, and conventional information criteria (namely AIC and BIC). The results are reported in Table 3. Under these measures of goodness of fit, two facts arise. First, the role of non-separation in the equilibrium correction terms is important to explain the variability of stock returns: columns 2 and 4 highlight the improvement in the in-sample predictive performance of the models when the futures bases from different stock markets are incorporated as explanatory variables in the returns equations. Second, the role of nonlinearities appears to be important to better explain stock returns. Columns 3 and 4 show how nonlinearities of the type specified in Section 2 help to capture the general features exhibited by the time series under investigation. However, examining the last column of Table 3, where international spillovers and nonlinearities are both explicitly taken into account, suggests that the in-sample performance of the model is very satisfactory. Even correcting for the larger number of parameters of the MSIAH(8)-VECM(1), the coefficient of determination of the latter model is at least four times larger than the coefficient of determination obtained for the bivariate

[19] It is worth noting that the regularity conditions under which the Davies (1977, 1987) test is valid may be violated, since the Markov model has both a problem of nuisance parameters and a problem of 'zero score' under the null hypothesis. Moreover, even if the Davies bound is appropriate, it is possible that it will only be valid if the null model is a linear model with iid errors; in the present case, it is difficult to believe that this condition is met since innovations are not homoskedastic, which would induce some distortion. Therefore, the distribution of the LR test may differ from the adjusted χ^2 distribution proposed by Davies (1977, 1987). For extensive discussions of the problems related to LR testing in this context, see Hansen (1992b, 1996) and Garcia (1998). We are thankful to Bruce Hansen for clarifying several econometric issues related to LR testing in the present context.

MS-VECMs and more than ten (five) times larger than the coefficient of determination of the standard linear VECM for the S&P 500 and the NIKKEI 225 (FTSE 100).[20]

The incremental explanatory power delivered by the MSIAH(8)-VECM(1) appears to be due to the fact that international spillovers and regime shifts are both important. Indeed, the lagged futures bases for each of the three markets considered are generally found to be strongly statistically significantly different from zero - i.e. with the equilibrium errors from the three markets found significant in other equilibrium correction models examined. As one might expect, although for each market stock returns respond most strongly to the futures basis in the domestic market, the US equilibrium error is found to be massively significant in each of the estimated nonlinear equilibrium correction models, suggesting that spillover effects from the S&P 500 may be particularly important. Moreover, for each stock market, the estimated equilibrium correction parameters display sizable shifts across regimes, clearly suggesting that equilibrium correction occurs in a nonlinear fashion. It seems, therefore, that the information contained in the equilibrium correction terms, both in the domestic market and the foreign markets, combined with the allowance for regime-switching behavior in the rich parametrization of the MSIAH(8)-VECM(1), is responsible for delivering the satisfactory explanatory power in terms of $\overline{R}^2$ and information criteria reported in Table 3.

4 Empirical analysis II: forecasting

4.1 Point forecasting performance and market timing tests

One of our results, corroborating some previous findings in the relevant literature, is that futures prices contain valuable information that can be exploited to explain a sizable proportion of stock prices and returns, at least in sample. In order to better evaluate the gain from using a sophisticated nonlinear empirical model, dynamic out-of-sample forecasts of stock returns were constructed using each of the models estimated and discussed in the previous section. In particular, we calculated one-step-ahead forecasts over the period January 1999-December 2002.[21] The out-of-sample forecasts are constructed according to a recursive procedure that is conditional only upon information available up to the date of the forecasts and with successive re-estimation as the date on which forecasts are conditioned moves through the data set. Also, note that given the definition of our equilibrium correction term z_t in equation (3), in estimation the constant of the cointegrating relationship is backed out per contract and is, in some sense, 'forward looking.' While this is the case in sample, however, our forecasting procedure does not use this information in that we estimate the constant recursively as we move through the forecast period.

It is well known in the literature that forecasting with nonlinear models is in general much more difficult than forecasting with linear models because of the need to condition on the distribution of future exogenous shocks whose conditional expectation may be zero in a linear framework but not in a nonlinear framework. However, given that we compute one-step-ahead

[20] Note that all of our estimated MSIAH-VECMs are stationary, as confirmed by calculating Karlsen's (1990) statistic. Moreover, as a way to evaluate the dynamic properties of the estimated Markov-switching models we also examined the effects of shocks on the evolution of the time series under investigation using generalized impulse response functions calculated using Monte Carlo integration methods (see Gallant, Rossi and Tauchen, 1993; Koop, Pesaran and Potter, 1996). The impulse response functions (not reported to conserve space but available upon request) show that, as expected, shocks hitting each of the three stock returns examined exhibit low persistence, dissipating over a very short time horizon.

[21] For a description of the econometric issues related to out-of-sample forecasting in a Markov-switching framework, see Hamilton (1993).

forecasts, the procedure often suggested in the literature that involves implementing numerical integration using Monte Carlo methods is not required as the one-step-ahead forecasts can be calculated analytically for our models (see, *inter alia*, Brown and Mariano, 1984, 1989; Granger and Teräsvirta, 1993, chapter 8; Franses and van Dijk, 2000, chapters 3-4; Krolzig, 2000).

Forecast accuracy is evaluated using several criteria. Panel a) of Table 4 shows the mean absolute error (MAE), the root mean square error (RMSE) and the out-of-sample R^2 for each of the estimated models. The out-of-sample R^2 is always higher for the MSIAH-VECM (14) than for any of the other competing models, resembling the adjusted R^2 obtained in sample. Indeed, the MSIAH-VECM (14) exhibits the best out-of-sample performance: the MAEs and RMSEs are always lower than the ones obtained from each of the alternative models suggesting that *both* nonlinearities *and* spillovers are important to explain, even out-of-sample, the dynamics of stock returns. However, the results of the Diebold-Mariano (DM, 1995) test, reported in parentheses in Panel a) of Table 4, indicate that we are not able to reject the null of equal predictive accuracy in each case. Hence the differences in terms of MAEs and RMSEs reported in Table 4 are not statistically significant and do not enable us to discriminate across the models examined.[22] [23] While no theoretical explanation exists for the similarity of the statistical performance of our linear and nonlinear models in terms of MAEs and RMSEs, this kind of finding has often been recorded in the relevant literature (e.g. Clements and Krolzig, 1998; Stock and Watson, 1999; Kilian and Taylor, 2003). One possibility is that the non-rejection of the null of equal point forecast accuracy under the DM test may be due to the low power of this test statistic in finite sample (e.g. Kilian and Taylor, 2003, and the references therein). This leads us to consider additional tests.

Alternative formal comparisons of the predicted and actual stock index returns can be obtained in a variety of ways. Hence, we consider the 'hit' rate, calculated as the proportion of correctly predicted signs of future stock price changes over the whole forecast period. Further, we consider a set of tests for market timing ability of the competing models. In particular, we carried out the tests proposed by Henriksson and Merton (1981), by Cumby and Modest (1987), and by Bossaerts and Hillion (1999) - HM, CM and BH tests from now onwards. The idea behind the HM test is that there is evidence of market timing if the sum of the estimated conditional probabilities of correct forecasts (that is the probability of correct forecast sign either when the market is bullish or bearish) exceeds unity. The HM test statistic is given by:

$$HM = \frac{n_{11} - \frac{n_{01}n_{10}}{n}}{\sqrt{\frac{n_{01}n_{10}n_{20}n_{02}}{n^2(n-1)}}} \sim N(0,1) \qquad (15)$$

where n_{11} is the number of correct bear market forecast; n_{01}, n_{10} are the number of bear markets and bear market forecasts respectively, while n_{02} and n_{20} denote the number of bull market and bull market forecasts respectively. The total number of evaluation periods is denoted by

[22] A consistent estimate of the spectral density at frequency zero $\widehat{f}(0)$ is obtained using the method of Newey and West (1987) where the optimal truncation lag has been selected using the Andrews's (1991) AR(1) rule. The rule is implemented as follows: we estimated an AR(1) model to the quantity d_t obtaining the autocorrelation coefficient $\widehat{\rho}$ and the innovation variance from the AR(1) process $\widehat{\sigma}^2$. Then the optimal truncation lag A for the Parzen window in the Newey and West estimator is given by the Andrews' rule $A = 2.6614\left[\widehat{\zeta}(2)T\right]^{1/5}$ where $\widehat{\zeta}(2)$ is a function of $\widehat{\rho}$ and $\widehat{\sigma}^2$. The Parzen window has been chosen according to Gallant (1987, p. 534).

[23] Note that the finite-sample distribution of the DM statistics may deviate from normality; this problem is particularly severe when one takes into account parameters uncertainty (see West 1996, West and McCracken 1998; McCracken 2000). The DM statistics reported in this paper were calculated by bootstrap (see Kilian, 1999).

n. The CM test extends the HM test to take into account not only the sign of the realized returns, but also their magnitude. This involves estimating the auxiliary regression:

$$\Delta s_{t+1}^i = \phi_0 + \phi_1 I_{\left\{\widetilde{\Delta s_{t+1}^i}>0\right\}} + error\ term, \tag{16}$$

where Δs_{t+1}^i is the time series of the realized returns for stock index i, and $I_{\left\{\widetilde{\Delta s_{t+1}^i}>0\right\}}$ is the indicator function equal to unity when the forecast returns for the index i, $\widetilde{\Delta s_{t+1}^i} > 0$ and is equal to zero otherwise. Finally, the BH test involves estimating the following auxiliary regression:

$$\Delta s_{t+1}^i = \zeta_0 + \zeta_1 \widetilde{\Delta s_{t+1}^i} + error\ term, \tag{17}$$

where $\widetilde{\Delta s_{t+1}^i}$ is the time series of the forecast returns for the stock index i. For both CM and BH tests, the null hypothesis of no market timing ability is that the slope coefficients ϕ_1 and ζ_1 are equal to zero against the one-sided alternative that they are positive. The results from executing these tests are reported in Panel b) of Table 4. Under these measures of market timing ability, we find a very different picture from the one suggested by the Diebold-Mariano tests for equal point forecast accuracy, but a similar picture to the one portrayed by the in-sample analysis. The role of non-separation in the equilibrium correction terms is important to explain out-of-sample futures and spot returns: columns 2 and 4 highlight the improvement in the predictive performance of the models when the futures bases from different stock markets are incorporated as explanatory variables in the returns equations. Thus, examining the last column of Table 4, where international spillovers and nonlinearities are both explicitly taken into account, suggests that the market timing performance of the MSIAH-VECM with international spillovers is highly satisfactory.

4.2 Density forecasting performance: main results

The findings in the previous subsection deserve further discussion. The estimated linear and nonlinear models produced a series of dynamic out-of-sample forecasts. Using different criteria to evaluate their predictive accuracy we obtained somewhat conflicting results. For example, the finding that the MSIAH-VECM with international spillovers displays satisfactory market timing ability relative to the various alternative models may seem at odds with its inability to beat the alternative models on the basis of MAEs and RMSEs. However, one explanation of these results is that, while the competing models are very similar in terms of their ability to forecast the conditional mean of stock returns, the MSIAH-VECM with international spillovers produces more accurate forecasts of the direction of future stock returns.

To shed further light on the forecasting ability of our models, we attempt to exploit the whole information provided by the MS-VECMs' out-of-sample predictions. In particular, the MSIAH-VECM (14) may exhibit the best performance across the models considered in terms of 'closeness' of the predicted moments to the true moments of stock returns data over the forecast period, although this might not be clear if one considers only the first two moments of the distribution of stock returns.

A logical next step then involves testing formally the hypothesis that the forecast density implied by the MSIAH-VECM (14) is equal to the true predictive density of the data. A large body of literature in financial econometrics has recently focused on evaluating the forecast accuracy of empirical models on the basis of density, as opposed to point, forecasting performance (see, *inter alia*, Diebold, Gunther and Tay, 1998; Diebold, Hahn and Tay, 1999; Granger and Pesaran, 1999; Tay and Wallis, 2000; Timmermann, 2000; Pesaran and Skouras, 2002; Sarno and Valente, 2003). Several researchers have proposed methods for evaluating density forecasts. For example, Diebold, Gunther and Tay (1998) extend previous work on the probability integral transform and show how it is possible to evaluate a model-based predictive density and to test formally the hypothesis that the predictive density implied by a particular model corresponds to the true predictive density. In general, they propose the calculation of the probability integral transforms of the actual realizations of the variables (i.e. stock returns for the different stock indices under investigation) over the forecast period, $\left\{\Delta s_{t+1}^{i}\right\}_{t=1}^{n}$ with respect to the models' forecast densities, denoted by $\left\{p_t\left(\Delta s_{t+1}^{i}\right)\right\}_{t=1}^{n}$:

$$w_t = \int_{-\infty}^{\Delta s_{t+1}^{i}} p_t\left(u\right) du \qquad t = 1, \ldots, n. \tag{18}$$

When the model forecast density corresponds to the true predictive density, denoted by $q_t\left(\Delta s_{t+1}^{i}\right)$, then the sequence of $\{w_t\}_{t=1}^{n}$ is iid $U\left[0,1\right]$. The idea is therefore to evaluate whether the realizations of the data over the forecast period do come from the selected forecast density by testing whether the $\{w_t\}$ series depart from the iid uniformity assumption. Following Clements and Smith (2000), we assess uniformity by plotting the empirical distribution function against the 45^0 line.[24] Berkowitz (2001) suggests that rather than working with the $\{w_t\}$ series it may be fruitful to take the inverse normal cumulative distribution function (CDF) transform of the series $\{w_t\}$, denoted by $\{x_t\}$. Under the null hypothesis of equality of the model density and the true predictive density, $\{x_t\}$ is distributed as standard normal, and Berkowitz proposes an LR test for zero mean, unit variance and independence. We rely on both the test of Diebold, Gunther and Tay (1998) and the test of Berkowitz (2001) in our empirical work.

While, under general conditions, the linear VECMs forecast densities are easy to calculate analytically (they are in fact multivariate normal distributions with means and variances given by simple functions of the estimated parameters), the implied MSIAH-VECM forecast densities can, in general, be obtained analytically only for one-step ahead forecasts. The MSIAH-VECM forecast densities are mixtures of multivariate normal distributions with weights given by the predicted regime probabilities. In general the MSIAH-VECM forecast densities are non-normal, asymmetric, heteroskedastic and regime dependent. Following Krolzig (2000), the one-step ahead MSIAH-VECM forecast density is given by:

$$p_{t+1}\left(\Delta y_{t+1}\right) = \sum_{j=1}^{M}\left\{\sum_{i=1}^{M} p_{ij}\mathbf{P}\right\} p_{t+1}\left(\Delta y_{t+1} \mid \omega_{t+1} = j, \Omega_t\right), \tag{19}$$

where $p_{ij} = \Pr(\omega_{t+1} = j \mid \omega_t = i)$ are the transition probabilities, $\mathbf{P}$ is the transition matrix conditional on the information set at time t, Ω_t and $p_{t+1}\left(\Delta y_{t+1} \mid \omega_{t+1} = j, \Omega_t\right)$ is the regime-conditional forecast density.

We now turn to the evaluation of the probability integral transforms. The null of iid uniformity is a joint hypothesis and, following the suggestion of Diebold, Gunther and Tay (1998), we consider each part of the hypothesis in turn. The iid assumption is tested by executing the Ljung-Box (1978) test for serial correlation up to the fourth order. The results are reported in Panel a) of Table 5. In order to take into account the dependence occurring in the higher moments, we also consider $(w - \overline{w})^j$ for j up to three. The results tell us that in

[24] It is important to notice that the confidence intervals reported are only valid under the assumption of independence.

most cases we are not able to reject the null hypothesis of no serial correlation. This finding applies particularly for the most general model, the MSIAH-VECM in equation (14), while some rejections are recorded in the second moment in the case of linear VECMs (11) and (12) estimated for FTSE 100 and S&P 500, and for the MSIAH-VECM (13) only for the FTSE 100. We assess the uniformity aspect by plotting the actual CDFs of the $\{w_t\}$ series against the theoretical CDF (i.e. 45^0 line), calculating the confidence intervals by Monte Carlo simulation with 50,000 replications. The results are plotted in Figure 2, which clearly indicates that it is possible to distinguish among the different competing models. In fact, for the models which consider either nonlinearity or international spillovers (but not both) we generally reject the null hypothesis of uniformity and in all cases we can see that the empirical CDFs exhibit an S-shape pattern around the 45^0 line. This could occur because the point forecast is a biased predictor of the mean of the true forecast density or, perhaps more likely in our context, it could be due to any of the higher moments failing to match. A different picture can be seen by looking at the last column in Figure 2. The most general model incorporating both nonlinearities and international spillovers does not exhibit the same S-shape pattern and, most importantly, we are not able to reject the null hypothesis of uniformity. Similar results can be found in Panel b) of Table 5, where we report the LR tests of zero mean, unit variance and independence proposed by Berkowitz (2001). In fact, the only model for which we cannot reject the null hypothesis is the MSIAH-VECM (14), with the exception of the NIKKEI 225 where we record a marginal non-rejection with a p-value of 0.048.

Summing up, the forecasting results in this section suggest that, in terms of density forecasting performance, the general MSIAH-VECM that allows for international spillovers performs better than any other linear and nonlinear model considered in this paper in terms of explaining the out-of-sample behavior of stock returns. These results should be taken with caution, however, since we are not directly testing one model against another, but comparing each model-based density to the true predictive density. This is because there is no test available to date which would allow us to make a direct comparison of competing models in terms of their density forecasting performance.[25] Taken together, the results in Section 4.1 and 4.2 suggest that, while the forecasting performance of the general MSIAH-VECM is not statistically different from the performance of the alternative models in terms of point forecasting, the MSIAH-VECM is superior when one evaluates out-of-sample performance on the basis of the ability of the model to match the full out-of-sample predictive density of stock returns. Clearly, this finding is due to the allowance for both international spillovers and multiple regimes in our model. This suggests that, although the various models examined do not differ statistically in terms of their predictive performance with respect to the conditional mean, the most general nonlinear model we propose provides a better characterization of the uncertainty surrounding the point forecasts.

4.3 The economic value of density forecasts: a simple example of Value-at-Risk analysis

Under the 1997 Amendment to the Basle Accord, banks may seek approval for the adoption of their own in-house risk models in order to calculate the minimum required capital to cover their market risk. Given that banks are permitted to develop different risk models, it is necessary to

[25] One test recently developed for the null hypothesis of equal density forecast accuracy is the nonparametric test developed by Sarno and Valente (2003). However, this test is not desirable in the present context since it assumes time-invariance of the predictive densities. We are grateful to an anonymous referee for useful comments on these issues.

assess the relative performance of the alternative models. Therefore it is interesting to further investigate the practical implications of the density forecasting results reported in the previous sub-section in the context of a simple risk management exercise. Given the predictions of the four competing models examined here, assume that a risk manager wishes to quantify the one-week-ahead risk associated with holding a stock index.[26] The different competing models provide the one-week-ahead density forecasts of Δs^i_{t+1} and on the basis of these densities the risk manager calculates the Value-at-Risk (VaR) of the stock index as a one-sided confidence interval on losses such that:

$$\Pr\left(\Delta s^i_{t+1} < VaR_{it+1}\right) = 1 - \phi, \tag{20}$$

where Δs^i_{t+1} is the realized end-of-week return and ϕ denotes the given confidence level. In our example the VaR is a 99 percent confidence level for losses (i.e. $\phi = 0.99$), for all models. Equation (20) simply states that the probability that the change in the value of the stock is less than the Value-at-Risk is equal to the significance level $1 - \phi$. Summary statistics are reported in Panel a) of Table 6. For all competing models we record the average VaR and the standard deviation of the estimated VaR over the forecast period and the realized violations, that are the number of times that $\Delta s^i_{t+1} < VaR_{it+1}$. The results in Table 6 suggest that, for all stock indices, the MSIAH-VECM (14) exhibits the highest average VaR (in absolute value), the highest standard deviation for the estimated VaR, and the lowest number of violations (i.e. zero). Although the latter result may suggest conservative behavior in predicting future risk, the high variability and the positive and significant correlation between the estimated VaR_{it+1} and the realized series of returns Δs^i_{t+1} are instead supportive of a fairly satisfactory performance of the MSIAH-VECM (14).[27]

In the literature, there is no definitive measure of VaR model performance. Thus, in order to evaluate the performance of the competing models, we present several different metrics. To assess the relative size and relative variability of the VaR estimates produced by the competing models we use the mean relative bias statistic (MRB) and root mean squared relative bias statistic (RMSRB), suggested by Hendricks (1996). The MRB statistic is calculated as:

$$MRB_i = \frac{1}{n}\sum_{j=1}^{n} \frac{VaR_{it+j} - \overline{VaR_{t+j}}}{VaR_{it+j}} \tag{21}$$

where VaR_{it} is the estimated Value-at-Risk for model i at time t, and $\overline{VaR_t}$ is the cross-sectional average (over the competing models) Value-at-Risk at time t. This statistic gives a measure of size for each estimated VaR relative to the average of all competing models. The RMSRB statistic is calculated as:

$$RSMRB_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{VaR_{it+j} - \overline{VaR_{t+j}}}{VaR_{it+j}}\right)^2}. \tag{22}$$

This measure provides us with information about the extent to which the estimated VaR tends to vary around the average VaR at time t. Another statistic, introduced by Christoffersen and Diebold (2000), is given by the first-order autocorrelation coefficient of a binary variable,

[26] Of course, a more complicated example would involve considering the joint density for all of the stock indices considered. We limit ourselves to the simplest case, given the illustrative nature of the application in the present section.

[27] The latter measure was introduced by Hendricks (1996, p. 161).

say V_t, which is equal to 1 if a violation occurs and 0 otherwise. A significant autocorrelation coefficient denotes a persistent series of violations which in turn implies unsatisfactory performance of a model in estimating the VaR.

The results, reported in Panel b) of Table 6, confirm the findings in Panel a). In fact the MRB and RSMRB statistics show that the MSIAH-VECM (14) produces lower VaRs (compared to the average VaR produced by all competing models) and it also produces more volatile VaRs (around the average VaR produced by all competing models). Since the MSIAH-VECMs (14) does not display violations over the forecast period, we are not able to calculate the Christoffersen-Diebold statistic for this model. For the remaining competing models, it is worth noting that the MSIAH-VECM (13) and the linear VECM (12) estimated for the FTSE 100 produced VaRs which experience persistent violations.

Summing up, this simple application further illustrates how the forecasting performance of alternative models can be very different when analyzed under different metrics. Conventional measures of predictive accuracy based on MAEs and RMSEs, recorded in previous sections, failed to recognize differences in higher moments of the predictive distributions. However, these features may be very relevant, for example, when assessing risk. In our example, although all the competing models were indistinguishable from the MSIAH-VECM (14) in terms of point forecast accuracy, they have produced forecasts that did not capture satisfactorily the higher moments of the predictive distribution of stock returns, generating VaRs that underestimate the probability of large losses. We find that the most general model incorporating both nonlinearities and international spillovers does not provide a perfect violation rate of unity, being consistent with a violation rate of zero. However, this more conservative model does better than all of the linear and nonlinear competing models at matching the higher moments of the predictive distribution of stock returns, generating VaRs that are more in line with the target violation rate of one percent.

5 Conclusion

This article has re-examined the dynamic relationship between spot and futures prices in stock index futures markets using data since 1989 at weekly frequency for three major stock market indices - the S&P 500, the NIKKEI 225 and the FTSE 100 indices. In particular, we propose a nonlinear, Markov-switching vector equilibrium correction model that explicitly takes into account the mounting evidence that the conditional distribution of stock returns is well characterized by a mixture of normal distributions. Also, we use the notion of 'separation and cointegration' to provide a richer characterization of the dynamics of stock returns that explicitly allows for international spillovers across these stock index and stock index futures markets.

The empirical results provide evidence in favor of the existence of international spillovers across these major stock markets and a well-defined long-run equilibrium relationship between spot and futures prices which is consistent with mean reversion in the futures basis. Linear vector equilibrium correction models were rejected when tested against a Markov-switching vector equilibrium correction model which allows for shifts in the intercept, the autoregressive structure and the variance-covariance matrix. Our preferred nonlinear specification explains a significant fraction of the stock returns examined, with the $\overline{R}^2$ ranging from 0.08 for the NIKKEI 225 index returns to 0.12 for the FTSE 100 index returns.

Using the estimated models in an out-of-sample forecasting exercise we found that both nonlinearity and international spillovers are important in forecasting stock returns. However, their importance is not apparent when the forecasting ability of our proposed nonlinear VECM is evaluated on the basis of conventional point forecasting criteria. In fact, these criteria neglect the fact that stock returns may be non-normally distributed and that the nonlinear models employed in this paper imply non-normal predictive densities. In order to measure more adequately the forecasting ability of our nonlinear model and discriminate among competing models we calculated hit rates, employed tests for market timing ability and evaluated the density forecasting performance of both linear and nonlinear models.

Overall, our empirical evidence suggests that the statistical performance of the linear (single-regime) and nonlinear (multiple-regime) models examined differs little in terms of conditional mean, regardless of whether allowance is made for international spillovers across the stock indices examined. However, calculation of hit rates and tests of market timing ability as well as inspection of predictive densities which fully consider the higher-order conditional moments implied by the various models show greater ability to discriminate between competing models. In particular, exploration of the model-based forecast densities indicates the rejection of single-regime models as well as multiple-regime models with no international spillovers against a multiple-regime model with international spillovers, leading us to the conclusion that both multiple regimes and the allowance for international spillovers are important ingredients for a model to produce satisfactory out-of-sample forecasting performance. The implication of our findings are further investigated in the context of a simple application to Value-at-Risk calculations which highlight how better density forecasts of stock returns, of the type recorded in this paper, can potentially lead to substantial improvements in risk management and, more precisely, to better estimates of downside risk.

While these results aid the profession's understanding of the behavior of stock returns, we view our model as a tentatively adequate characterization of the data which appears to be superior to linear equilibrium correction modeling in a number of respects, but which nevertheless may be capable of improvement. In particular, while we focused on the information provided by the futures market for forecasting stock returns, it would be interesting to investigate the presence of regime-switching behavior in the context of conventional models involving dividend yields or other fundamentals. Also, while the model used here is fairly general and flexible, the evidence we document suggests that global stock index and stock index futures markets are characterized by very complex dynamic interactions. Much more work needs to be done to understand these relationships.

Table 1. Preliminary data statistics

	S&P 500		NIKKEI 225		FTSE 100	
	f_t	s_t	f_t	s_t	f_t	s_t
Minimum	7.333	7.318	10.589	10.569	8.837	8.829
Maximum	5.645	5.632	9.025	9.029	7.506	7.491
Mean	6.489	6.483	9.833	9.829	8.227	8.220
Std Dev	0.521	0.520	0.321	0.317	0.375	0.378
PACF:						
lag 1	0.997	0.997	0.990	0.990	0.995	0.996
lag 2	0.029	0.030	-0.024	-0.015	0.066	0.052
lag 3	0.010	0.003	0.002	-0.006	0.016	0.008
lag 4	-0.013	-0.010	0.000	-0.006	-0.002	-0.012
lag 5	0.013	0.019	0.003	0.007	0.056	0.069
lag 6	0.018	0.016	-0.044	-0.044	0.035	0.028
lag 7	-0.017	-0.016	0.013	0.015	-0.023	-0.024
lag 8	0.011	0.012	0.012	0.013	0.011	0.010
lag 9	0.024	0.012	-0.032	-0.029	0.026	0.009
lag 10	-0.011	-0.009	0.015	0.012	0.008	0.016
lag 11	0.001	0.002	-0.005	0.001	0.026	0.014
lag 12	-0.006	-0.006	-0.011	-0.018	-0.033	-0.040

Notes: f_t and s_t denote the log-level of the futures price and the log-level of the spot price respectively. PACF is the partial autocorrelation function, and its standard deviation can be approximated by the square root of the reciprocal of the number of observations.

Table 2. Specification tests results

	LR1	J-test	LR2
Bivariate VECM for S&P 500	359.44	21.01	—
	$\{8.00 \times 10^{-69}\}$	$\{3.15 \times 10^{-4}\}$	—
Bivariate VECM for NIKKEI 225	270.10	13.09	—
	$\{4.22 \times 10^{-50}\}$	$\{1.08 \times 10^{-2}\}$	—
Bivariate VECM for FTSE 100	340.60	11.08	—
	$\{7.35 \times 10^{-65}\}$	$\{2.57 \times 10^{-2}\}$	—
Multivariate VECM (all indices)	1999.56	2517.52	1698.30
	$\{7.09 \times 10^{-144}\}$	$\{0\}$	$\{1.72 \times 10^{-98}\}$

Notes: $LR1$ tests the null hypothesis of a linear VECM against the alternative hypothesis of an MSIAH-VECM with $M = 2$ or 2^3 regimes with lag-length $p = 5$. J-test is the Davidson and MacKinnon (1981) test for non-nested models. The figures reported are relative to the test of $\Psi = \mathbf{0}$ in the auxiliary regression $\Delta y_t = (\mathbf{I} - \Psi)\,\Delta\widehat{y}_t^L + \Psi\Delta\widehat{y}_t^{NL}$, where $\Delta y_t = [\Delta f_t, \Delta s_t]'$ and $\Delta\widehat{y}_t^L$, $\Delta\widehat{y}_t^{NL}$ are the predictions from VECMs and MSIAH-VECMs respectively. $LR2$ is the likelihood ratio test calculated to test the restrictions in (9) for the estimated MSIAH-VECMs. The tests statistics are distributed as $\chi^2(g)$ where g is the number of restrictions imposed. Figures in braces denote p-values. For LR1 p-values are calculated as in Davies (1977, 1987).

Table 3. In-sample performance

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		$\overline{R}^2$		
S&P 500	0.006	0.023	0.026	0.101
NIKKEI 225	0.001	0.007	0.005	0.081
FTSE 100	0.019	0.029	0.027	0.118
		Information Criteria		
AIC	0.938	0.964	0.979	–
BIC	0.878	0.897	0.917	–

Notes: $\overline{R}^2$ is the adjusted coefficient of determination calculated for the stock returns equation in each VECM as in Krolzig (1997). AIC and BIC are the ratios of the AIC and the BIC from the MSIAH-VECM (14) to the corresponding goodness-of-fit measure for each of the alternative competing models. AIC and BIC criteria reported are calculated for the whole (linear and nonlinear) VECM systems.

Table 4. Out-of-sample performance: point forecasting

Panel a) Mean absolute errors, root mean square errors and Diebold-Mariano tests

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		S&P 500		
MAE	0.022 (0.923)	0.022 (0.927)	0.021 (0.925)	0.015 (−)
$RMSE$	0.028 (0.995)	0.028 (0.996)	0.027 (0.996)	0.019 (−)
R^2	0.010	0.019	0.038	0.123
		NIKKEI 225		
MAE	0.026 (0.933)	0.025 (0.939)	0.025 (0.938)	0.020 (−)
$RMSE$	0.033 (0.995)	0.032 (0.995)	0.032 (0.995)	0.024 (−)
R^2	0.005	0.017	0.008	0.086
		FTSE 100		
MAE	0.021 (0.922)	0.021 (0.923)	0.021 (0.921)	0.012 (−)
$RMSE$	0.027 (0.996)	0.027 (0.996)	0.027 (0.996)	0.015 (−)
R^2	0.015	0.017	0.036	0.106

Panel b) Market timing test

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		S&P 500		
HR	0.510	0.553	0.563	0.745
HM	3.54×10^{-1}	4.02×10^{-2}	1.19×10^{-3}	1.05×10^{-13}
CM	1.41×10^{-1}	2.46×10^{-2}	1.24×10^{-7}	8.88×10^{-20}
BH	3.84×10^{-2}	1.09×10^{-5}	7.61×10^{-11}	2.89×10^{-42}
		NIKKEI 225		
HR	0.534	0.601	0.519	0.697
HM	9.38×10^{-1}	6.85×10^{-3}	9.60×10^{-1}	6.65×10^{-8}
CM	2.51×10^{-1}	1.44×10^{-1}	1.68×10^{-1}	9.49×10^{-11}
BH	4.62×10^{-1}	7.89×10^{-2}	4.79×10^{-3}	5.01×10^{-29}
		FTSE 100		
HR	0.529	0.587	0.529	0.760
HM	2.30×10^{-1}	4.23×10^{-3}	2.30×10^{-1}	5.48×10^{-14}
CM	3.39×10^{-2}	5.43×10^{-4}	9.47×10^{-3}	1.39×10^{-21}
BH	5.54×10^{-7}	7.73×10^{-8}	4.59×10^{-9}	3.13×10^{-64}

Notes: *Panel a):* MAE and $RMSE$ denote the mean absolute error and the root mean square error respectively. Figures in parentheses are p-values from executing Diebold-Mariano (1995) test statistics for the null hypothesis that model i = VECM(11), VECM(12), MSIAH-VECM(13) have equal point forecast accuracy to the MSIAH-VECM(14). The spectral density of the loss differential function at frequency zero $\widehat{f}(0)$ is estimated using the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The p-values are calculated by bootstrap methods using a variant of the procedure suggested by Kilian (1999). R^2 is the out-of-sample coefficient of determination. *Panel b):* HR is the hit rate calculated as the proportion of correctly predicted signs. HM is the Henriksson and Merton (1981) test for the null of no market timing. CM is the Cumby and Modest (1987) test for the significance of the t-statistics of the slope coefficient in the regression $\Delta s_t^i = \phi_0 + \phi_1 I_{\left\{\widetilde{\Delta s_t^i}>0\right\}} + \epsilon$ where Δs_t^i are the realized returns for the index $i = S\&P\ 500, NIKKEI\ 225, FTSE\ 100$, and I is the indicator function equal to 1 when the forecast returns for the index i $\widetilde{\Delta s_t^i} > 0$ and equal to zero otherwise. BH is the Bossaerts and Hillion (1999) test for the significance of the t-statistics of the slope coefficient in the regression $\Delta s_t^i = \zeta_0 + \zeta_1 \widetilde{\Delta s_t^i} + \epsilon$ where $\widetilde{\Delta s_t^i}$ are the forecast returns for the index $i = S\&P\ 500, NIKKEI\ 225, FTSE\ 100$. For each of the HM, CM, and BH test statistics only p-values are reported.

Table 5. Out-of-sample performance: density forecasting

Panel a) Test for iid based upon probability integral transforms

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		S&P 500		
w	0.457	0.330	0.201	0.423
$(w-\overline{w})^2$	0.052	0.031	0.350	0.934
$(w-\overline{w})^3$	0.441	0.307	0.337	0.270
		NIKKEI 225		
w	0.501	0.505	0.417	0.968
$(w-\overline{w})^2$	0.957	0.934	0.489	0.175
$(w-\overline{w})^3$	0.507	0.436	0.333	0.477
		FTSE 100		
w	0.221	0.278	0.278	0.403
$(w-\overline{w})^2$	0.016	0.007	0.017	0.413
$(w-\overline{w})^3$	0.093	0.121	0.167	0.275

Panel b) Berkowitz (1999) LR test

VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
	S&P 500		
2.73×10^{-19}	1.04×10^{-19}	7.34×10^{-17}	1.18×10^{-1}
	NIKKEI 225		
3.86×10^{-3}	5.11×10^{-3}	2.78×10^{-9}	4.77×10^{-2}
	FTSE 100		
7.08×10^{-11}	6.12×10^{-13}	1.80×10^{-12}	7.59×10^{-2}

Notes: *Panel a):* Figures denote p-values for the Ljung and Box (1978) χ^2 test of serial correlation up to fourth order. *Panel b):* Figures denote p-values for the LR test of Berkowitz (2001). The tests are calculated considering an alternative model with quadratic and cubic terms lagged up to order 4. The test statistic is distributed under the null as a $\chi^2(l)$ where l is the number of restrictions imposed.

Table 6. Value-at-Risk exercise

Panel a) Summary statistics

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		S&P 500		
Mean VaR	-0.053	-0.052	-0.051	-0.058
S.D. VaR	8.39×10^{-6}	1.54×10^{-5}	5.58×10^{-5}	4.16×10^{-4}
n. violations	5	5	8	0
$corr\,(\Delta s_t, VaR)$	0.117	0.303*	0.198*	0.731*
		NIKKEI 225		
Mean Var	-0.078	-0.078	-0.070	-0.085
S.D. VaR	5.38×10^{-6}	1.86×10^{-5}	5.19×10^{-5}	5.92×10^{-4}
n. violation	5	4	7	0
$corr\,(\Delta s_t, VaR)$	-0.046	0.125	0.048	0.621*
		FTSE 100		
Mean VaR	-0.054	-0.054	-0.051	-0.061
S.D. VaR	1.17×10^{-5}	1.35×10^{-5}	2.85×10^{-5}	5.75×10^{-4}
n. violation	5	5	8	0
$corr\,(\Delta s_t, VaR)$	0.290*	0.294*	0.226*	0.814*

Panel b) VaR backtests

	VECM (11)	VECM (12)	MSIAH-VECM (13)	MSIAH-VECM (14)
		S&P 500		
MRB	-0.001	-0.011	-0.044	0.056
RMSRB	0.106	0.096	0.124	0.265
CD	-0.025	-0.025	-0.040	–
		NIKKEI 225		
MRB	0.010	0.013	-0.093	0.074
RMSRB	0.083	0.083	0.138	0.229
CD	-0.024	-0.020	-0.035	–
		FTSE 100		
MRB	-0.002	-0.007	-0.067	0.076
RMSRB	0.109	0.106	0.128	0.308
CD	-0.024	0.180*	0.219*	–

Notes: *Panel a):* "Mean Var" and "S.D. VaR" denote the mean and standard deviation of the calculated VaRs from model (11)-(14) respectively; "n. of violations" denotes the number of times when the realized returns exceeds the estimated VaR. $corr\,(\Delta s_t, VaR)$ is the correlation coefficient between the estimated VaR and the realized data, calculated as in Hendricks (1996). *Panel b):* MRB and RMSRB denote the mean relative bias and square-root mean relative bias respectively, calculated as in Hendricks (1996). CD is the Christoffersen and Diebold (2000) test for the sample autocorrelation of a binary variable which is equal to 1 if a violation occurs and zero otherwise. * denote significant statistics at the 1% level respectively.

A Appendix: The transition matrix of the MSIAH-VECM

In Section 2.2 we mentioned that the underlying regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $\omega_t \in \{1, \ldots, M\}$ governed by the transition probabilities $p_{ij} = \Pr(\omega_t = j \mid \omega_{t-1} = i)$, and $\sum_{j=1}^{M} p_{ij} = 1 \; \forall i, j \in \{1, \ldots, M\}$. If we move from the perspective of a single system of variables (i.e. futures and spot returns in a single stock market) towards a model where several systems of variables are jointly considered (i.e. non-separation is explicitly considered, MSIAH-VECM (14)), we need to specify the joint process governing the transitional dynamics of the whole system. Define ω_t^{SP}, ω_t^{NK} and ω_t^{FT} the unobserved variable governing the transitional dynamics of the S&P 500, NIKKEI 255 and FTSE 100 indices respectively, and assume $M = 2$.

In order to achieve greater flexibility, at the cost of a high computational burden, we make no assumption about the relationship between the shifts occurring in the three markets examined, so that ω_t^{υ} would be an outcome of a Markov chain with transition probabilities p_{ij}^{υ} where ω_t^{υ} is independent of ω_t^{ψ} with $\psi \neq \upsilon$ for any t. In order to analyze the whole dynamics of the MSIAH-VECM (14) we construct the following latent variable

$$\begin{array}{lcll}
\xi_t & = & 1 & \text{if } \omega_t^{SP} = 1, \omega_t^{NK} = 1 \text{ and } \omega_t^{FT} = 1 \\
\xi_t & = & 2 & \text{if } \omega_t^{SP} = 2, \omega_t^{NK} = 1 \text{ and } \omega_t^{FT} = 1 \\
\xi_t & = & 3 & \text{if } \omega_t^{SP} = 1, \omega_t^{NK} = 2 \text{ and } \omega_t^{FT} = 1 \\
\xi_t & = & 4 & \text{if } \omega_t^{SP} = 2, \omega_t^{NK} = 2 \text{ and } \omega_t^{FT} = 1 \\
\xi_t & = & 5 & \text{if } \omega_t^{SP} = 1, \omega_t^{NK} = 1 \text{ and } \omega_t^{FT} = 2 \\
\xi_t & = & 6 & \text{if } \omega_t^{SP} = 2, \omega_t^{NK} = 1 \text{ and } \omega_t^{FT} = 2 \\
\xi_t & = & 7 & \text{if } \omega_t^{SP} = 1, \omega_t^{NK} = 2 \text{ and } \omega_t^{FT} = 2 \\
\xi_t & = & 8 & \text{if } \omega_t^{SP} = 2, \omega_t^{NK} = 2 \text{ and } \omega_t^{FT} = 2.
\end{array} \tag{A1}$$

Under this formalization the latent variable ξ_t governing the transitional dynamics of the whole system MSIAH-VECM (14) follows an 8-state Markov chain whose transition probabilities can be easily calculated from the probabilities of the chain governing ω_t^{SP}, ω_t^{NK} and ω_t^{FT}. For example:

$$\begin{aligned}
\Pr\left(\xi_t = 1 | \xi_{t-1} = 1\right) &= \Pr\left(\omega_t^{SP} = 1 | \omega_{t-1}^{SP} = 1\right) \cdot \Pr\left(\omega_t^{NK} = 1 | \omega_{t-1}^{NK} = 1\right) \cdot \\
&\quad \Pr\left(\omega_t^{FT} = 1 | \omega_{t-1}^{FT} = 1\right) \\
&= p_{11}^{SP} p_{11}^{NK} p_{11}^{FT}.
\end{aligned} \tag{A2}$$

References

Ahn, D.H., Boudoukh, J., Richardson, M. and Whitelaw, R.F. (2002), "Partial Adjustment or State Prices? Implications from Stock Index and Futures Return Autocorrelation," *Review of Financial Studies*, 15, 655-689.

Andreou, E., Osborn, D.R. and Sensier, M. (2000), "A Comparison of the Statistical Properties of Financial Variables in the USA, UK and Germany over the Business Cycle," *Manchester School*, 68, 396-418.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Ang, A. and Bekaert, G. (2001), "Stock Return Predictability: Is it There?", *National Bureau of Economic Research Working Paper* No. 8207.

Aslanidis, N., Osborn, D.R. and Sensier, M. (2002), "Smooth Transition Regression Models in UK Stock Returns," University of Manchester, mimeo.

Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management," *Review of Economics and Statistics*, 19, 465-474.

Bossaerts, P. and Hillion, P. (1999), "Implementing Statistical Criteria to Select Returns Forecasting Models: What Do We Learn?", *Review of Financial Studies*, 12, 405-428.

Brock, W.A., Dechert, W.D. and Scheinkman, J. (1991), *Nonlinear Dynamics, Chaos, and Instability.* Cambridge, Mass.: MIT Press.

Cecchetti, S.G., Lam, P. and Mark, N.C. (1990), "Mean Reversion in Equilibrium Asset Prices," *American Economic Review*, 80, 398-418.

Cecchetti, S.G., Lam, P. and Mark, N.C. (2000), "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to be True?", *American Economic Review*, 90, 787-805.

Christoffersen, P.F. (1998), "Evaluating Interval Forecasting," *International Economic Review*, 39, 841-862.

Christoffersen, P.F. and Diebold, F.X. (2000), "How Relevant is Volatility Forecasting for Financial Risk Management?", *Review of Economics and Statistics*, 82, 12-22.

Clements, M.P. and Hendry, D.F. (1996), "Intercept Corrections and Structural Change," *Journal of Applied Econometrics*, 11, 475-494.

Clements, M.P. and Hendry, D.F. (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Clements, M.P. and Krolzig, H.-M. (1998), "A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP," *Econometrics Journal*, 1, C47-75.

Clements, M.P. and Smith, J. (2000), "Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 19, 255-276.

Cumby, R.E. and Modest, D.M. (1987), "Testing for Market Timing Ability, A Framework for Forecast Evaluation," *Journal of Financial Economics*, 19, 169-189.

Davies, R.B. (1977), "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 64, 247-254.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum Likelihood Estimation from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, 39, Series B, 1-38.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Diebold, F.X., Hahn, J. and Tay, A.S. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Dwyer, G.P. Jr., Locke, P.R. and Yu, W. (1996), "Index Arbitrage and Nonlinear Dynamics between the S&P 500 Futures and Cash," *Review of Financial Studies*, 9, 301-32.

Engle, R.F. and Granger, C.W.J. (1987), "Co-integration and Error Correction Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

Engle, R.F. and Susmel, R. (1994), "Hourly Volatility Spillovers between International Equity Markets," *Journal of International Money and Finance,* 13, 3-25.

Eun, C.S. and Shin, S. (1989), "International Transmission of Stock Market Movements," *Journal of Financial and Quantitative Analysis*, 24, 241-256.

Franses, P.H. and van Dijk, D. (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge and New York: Cambridge University Press.

French, K.R. (1980), "Stock returns and the Weekend Effect," *Journal of Financial Economics*, 8, 55-70.

Gagnon, L. and Lypny, G. (1995), "Hedging Short-term Interest Risk Under Time-varying Distributions," *Journal of Futures Markets*, 15, 767-783.

Gallant A.R (1987), *Nonlinear Statistical Models*, New York, NY: J. Wiley.

Gallant, A.R., Rossi, P.E. and Tauchen, G., (1993), "Nonlinear Dynamic Structures," *Econometrica,* 61, 871-907.

Gao, A.H. and Wang, G.H.K. (1999), "Modeling Nonlinear Dynamics of Daily Futures Price Changes," *Journal of Futures Markets*, 19, 325-351.

Garcia, R. (1998), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763-788.

Gibbons, M.R. and Hess, P.J. (1981), "Day of the Week Effects and Asset Returns," *Journal of Business*, 4, 579-596.

Goetzmann, W.N., Li, L. and Rouwenhorst, K.G. (2001), "Long-term Global Market Correlations," National Bureau of Economic Research Working Paper No. 8612.

Granger, C.W.J. (1986), "Developments in the Study of Cointegrated Variables," *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

Granger, C.W.I and Pesaran, M.H. (1999), "A Decision Theoretic Approach to Forecast Evaluation," in Chan, W.S., Lin, W.K. and Tong, H. (eds.), *Statistics and Finance: An Interface*, London: Imperial College Press.

Granger, C.W.J. and Swanson, N. (1996), "Future Developments in the Study of Cointegrated Variables," *Oxford Bulletin of Economics and Statistics,* 58, 537-53.

Granger, C.W.J. and Haldrup, N. (1997), "Separation in Cointegrated Systems and Persistent-Transitory Decompositions," *Oxford Bulletin of Economics and Statistics,* 59, 449-63.

Granger, C.W.I. and Pesaran, M.H., (1999), "A Decision Theoretic Approach to Forecast Evaluation," in Chan, W.S., Lin, W.K. and Tong, H. (eds.), *Statistics and Finance: An Interface*, London: Imperial College Press.

Granger, C.W.J. and Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Hamilton, J.D. (1988), "Rational Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates," *Journal of Economics Dynamics and Control*, 12, 385-423.

Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Hamilton, J.D. (1993), "Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime," in Maddala, G.S., Rao, C.R. and Vinod, H.D. (eds.), *Handbook of Econometrics* , Vol. 4, Elsevier, Amsterdam.

Hamilton, J.D. and Lin, G. (1996), "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 11, 573-93.

Hamilton, J.D. and Susmel, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307-333.

Hansen, B.E., (1992a) "Heteroskedastic Cointegration," *Journal of Econometrics*, 54, 139-158.

Hansen, B.E. (1992b), "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S, 61-82.

Hansen, B.E. (1996), "Erratum: The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 11, 195-198.

Hendricks, D. (1996), "Evaluation of Value-at-Risk Model Using Historical Data," *Federal Reserve Bank of New York Economic Policy Review*, April, 29-69.

Henriksson, R.D. and Merton, R.C. (1981), "On Market Timing and Investment Performance II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, 513-533.

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Jorion, P. (2000), *Value at Risk: The New Benchmark for Managing Financial Risk*, London and New York: McGraw-Hill.

Karlsen, H.A. (1990), *A Class of Non-Linear Time Series Models*, University of Bergen, unpublished Ph.D. dissertation.

Karolyi, G.A. and Stulz, R.M. (1996), "Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Comovements," *Journal of Finance* 51, 951-86.

Kendall, M.G. and Stuart, A. (1976), *The Advanced Theory of Statistics*, Vol. 1, 4th edition, London: Charles Griffin and Co.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?", *Journal of Applied Econometrics*, 14, 491-510.

Kilian, L. and Taylor, M.P. (2003), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?", *Journal of International Economics*, 60, 85-107.

Kim, C.-J. and Nelson, C.R. (1999), *State-Space Models with Regime Switching*, Cambridge, Mass. and London, UK: MIT Press.

Konishi, T. and Granger, C.W.J. (1993), "Separation in Cointegrated Systems," University of California San Diego mimeo.

Koop, G., Pesaran, H.M. and Potter, S. (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 77-118.

Koutmos, G. and Booth, G.G. (1995), "Asymmetric Volatility Transmission in International Stock Markets," *Journal of International Money and Finance*, 14, 747-762.

Krolzig, H.-M. (1997), *Markov-Switching Vector Autoregressions*, New York, NY: Springer.

Krolzig, H.-M. (1999), "Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts," Nuffield College and Department of Economics, University of Oxford, mimeo.

Krolzig, H.-M. (2000), "Predicting Markov-Switching Vector Autoregressive Processes", Nuffield College and Department of Economics, University of Oxford, mimeo.

Kroner, K.F. and Sultan, J. (1993), "Time-varying Distributions and Dynamic Hedging with Foreign Currency Futures," *Journal of Financial and Quantitative Analysis*, 28, 535-551.

LeBaron, B. (1992), "Forecast Improvements Using a Volatility Index," *Journal of Applied Econometrics*, 7, 137-49.

Lee, B.-S. and Jeon, B.N. (1995), "Common Stochastic Trends and Predictability of International Stock Prices," *Journal of the Japanese and International Economies*, 9, 245-77.

Lien, D. and Lou, X. (1993) "Estimating Multiperiod Hedge Ratios in Cointegrated Markets," *Journal of Futures Markets*, 13, 909-920.

Lien, D. and Lou, X. (1994) "Multiperiod Hedging in Presence of Conditional Heteroskedasticity," *Journal of Futures Markets*, 14, 927-955.

Ljung, G. and Box G. (1979) "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 265–270.

Low, A., Muthuswamy, J., Sakar, S. and Terry, E. (2002), "Multiperiod Hedging with Futures Contracts," *Journal of Futures Markets*, 22, 1179-1203.

Martens, M., Kofman, P. and Vorst, T.C.F. (1998), "A Threshold Error-Correction Model for Intraday Futures and Index Returns," *Journal of Applied Econometrics*, 13, 245-63.

Martens, M. and Poon, S.-H. (2001), "Returns Synchronization and Daily Correlation Dynamics Between International Stock Markets," *Journal of Banking and Finance*, 25, 1805-27.

McCracken, M.W. (2000), "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, 195-223.

Miller, M.H., Muthuswamy, J. and Whaley, R.E. (1994), "Mean Reversion of Standard & Poor's 500 Index Basis Changes: Arbitrage-Induced or Statistical Illusion?", *Journal of Finance*, 49, 479-513.

Neely, C.J. and Weller, P. (2000), "Predictability in International Asset Returns: A Reexamination," *Journal of Financial and Quantitative Analysis*, 35, 601-620.

Newey, W.K. and West, K.D. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 3, 703-708.

Osterwald-Lenum, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

Pan, M.S. and Hsueh, L.P. (1998), "Transmission of Stock Returns and Volatility Between the US and Japan: Evidence from the Stock Index Futures Market," *Asia-Pacific Financial Markets*, 5, 211-225.

Perez-Quiros, G. and Timmermann, A. (2001), "Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities," *Journal of Econometrics*, 103, 259-306.

Pesaran, M.H. and Skouras, S. (2002), "Decision-based Methods for Forecast Evaluation," in Clements, M.P. and Hendry, D.F. (eds.), *A Companion to Economic Forecasting*, Chapter 11, 241-267.

Pesaran, M.H. and Timmermann, A. (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201-28.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Least-Squares Regression Analysis," *Journal of the Royal Statistical Analysis, Series B*, 31, 350-371.

Ramchand, L. and Susmel, R. (1998a), "Volatility and Cross Correlation across Major Stock Markets," *Journal of Empirical Finance,* 5, 397-416.

Ramchand, L. and Susmel, R. (1998b), "Variances and Covariances of International Stock Returns: The International Capital Asset Pricing Model Revisited," *Journal of International*

Financial Markets, Institutions and Money, 8, 39-57.

Rydén, T., Teräsvirta, T. and Åsbrink, S. (1998), "Stylized Facts of Daily Return Series and the Hidden Markov Model," *Journal of Applied Econometrics,* 13, 217-44.

Saikkonen, P. (1992), "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1-27.

Saikkonen, P. and Luukkonen, R. (1997), "Testing Cointegration in Infinite Order Vector Autoregressive Processes," *Journal of Econometrics*, 81, 93-126.

Sarno, L. and Valente, G. (2003),"Comparing the Accuracy of Density Forecasts from Competing Models," *Journal of Forecasting*, forthcoming.

Stock, J.H. and Watson, M.W. (1999), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in Engle, R. and White, H. (eds.), *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, 1-44.

Susmel, R. (1999), "Switching Volatility in International Equity Markets," University of Houston, mimeo.

Tay, A.S. and Wallis, K.F. (2000), "Density Forecasting: A Survey," *Journal of Forecasting*, 19, 235-254.

Timmermann, A. (2000), "Density Forecasting in Economics and Finance: Editorial," *Journal of Forecasting*, 19, pp. 231-234.

Timmermann, A. (2001), "Structural Breaks, Incomplete Information, and Stock Prices," *Journal of Business and Economic Statistics*, 19, 299-314.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

West, K.D. and McCracken, M. W. (1998), "Regression-based Tests of Predictive Ability," *International Economic Review*, 39, 817-840.

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

Figure 1. Weekly Log-differences of Stock Prices

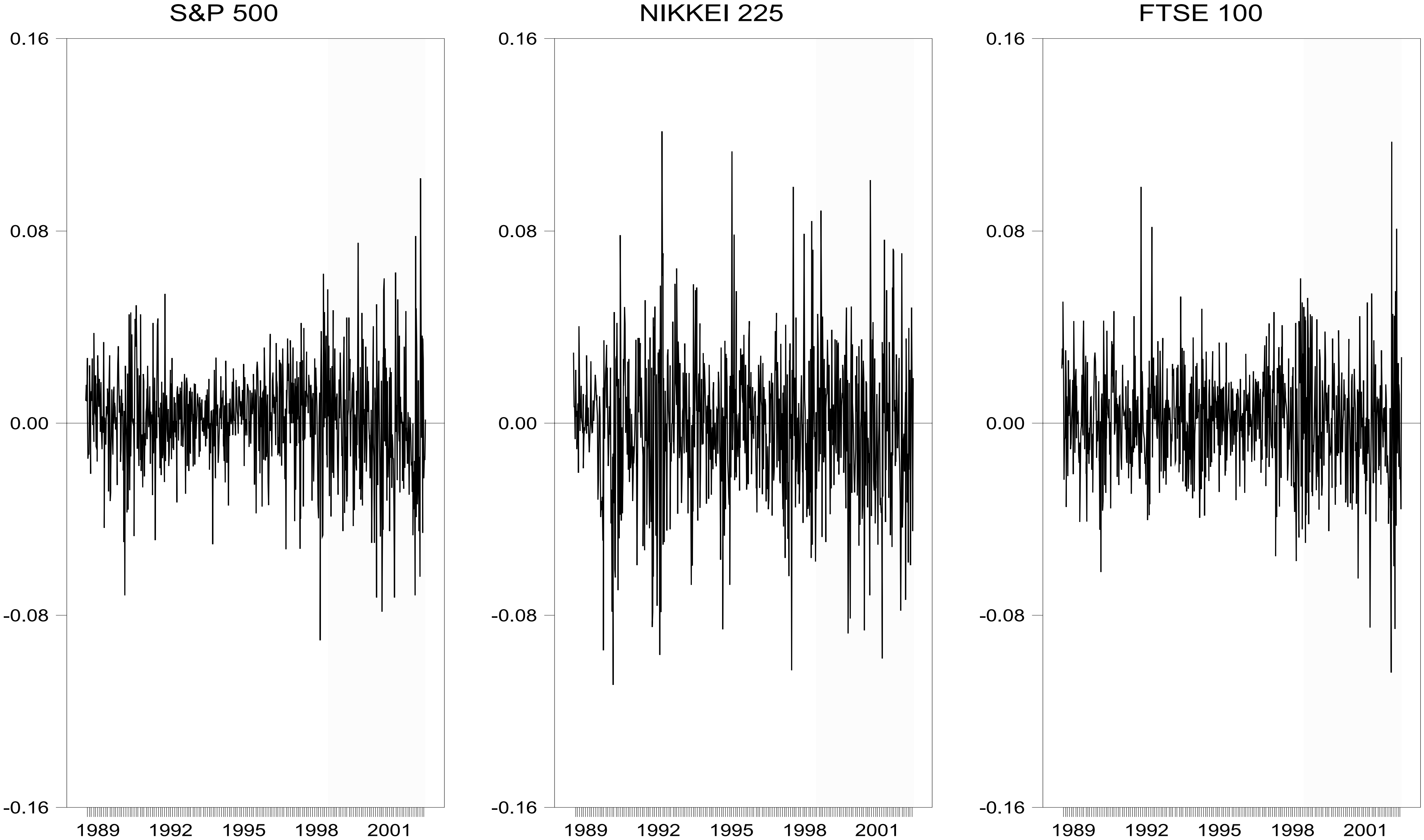
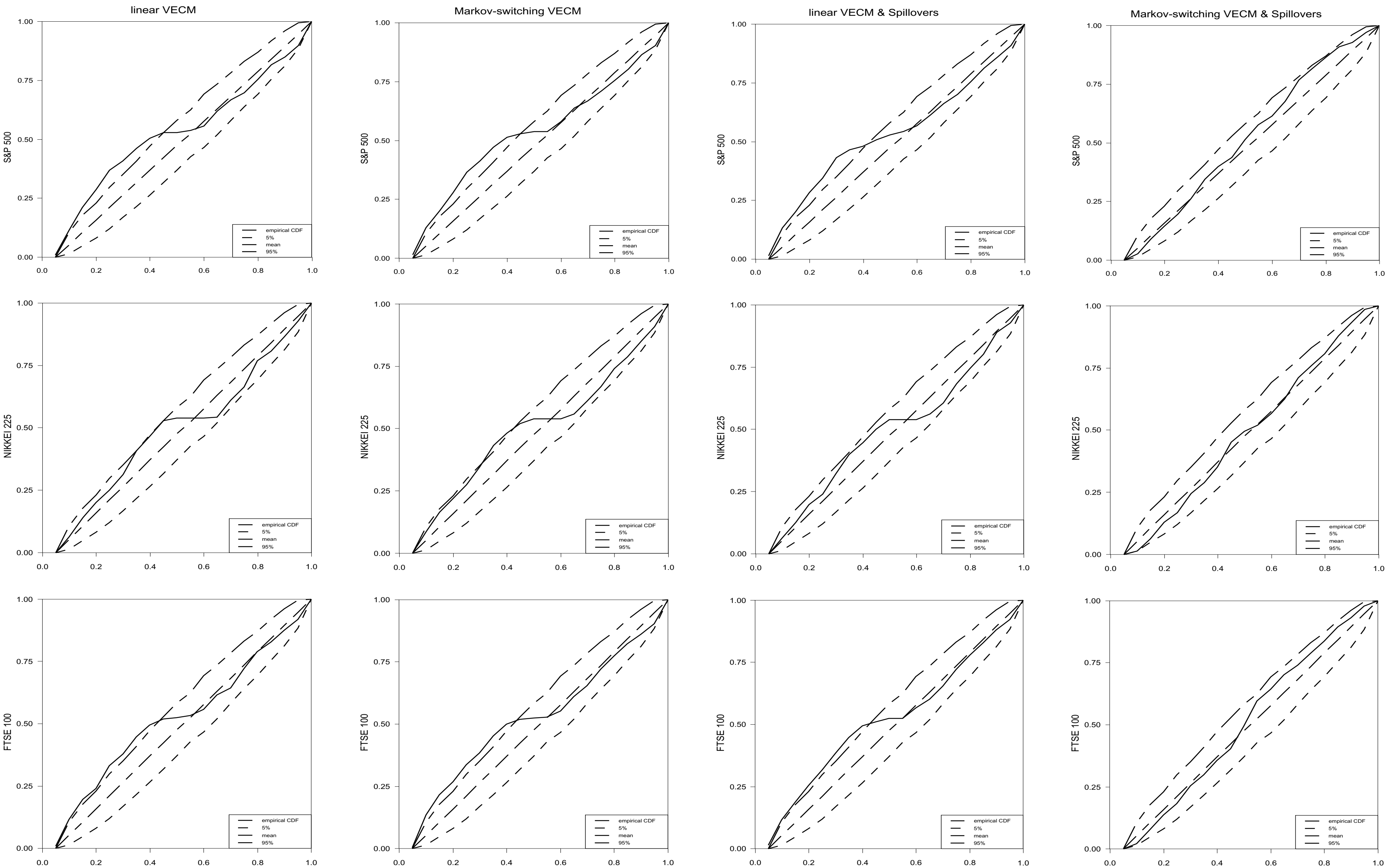


Figure 2. CDFs of w-values



List of other working papers:

2004

1. Xiaohong Chen, Yanqin Fan and Andrew Patton, Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates, WP04-19
2. Valentina Corradi and Walter Distaso, Testing for One-Factor Models versus Stochastic Volatility Models, WP04-18
3. Valentina Corradi and Walter Distaso, Estimating and Testing Sochastic Volatility Models using Realized Measures, WP04-17
4. Valentina Corradi and Norman Swanson, Predictive Density Accuracy Tests, WP04-16
5. Roel Oomen, Properties of Bias Corrected Realized Variance Under Alternative Sampling Schemes, WP04-15
6. Roel Oomen, Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling, WP04-14
7. Richard Clarida, Lucio Sarno, Mark Taylor and Giorgio Valente, The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates, WP04-13
8. Lucio Sarno, Daniel Thornton and Giorgio Valente, Federal Funds Rate Prediction, WP04-12
9. Lucio Sarno and Giorgio Valente, Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers, WP04-11
10. Lucio Sarno and Giorgio Valente, Empirical Exchange Rate Models and Currency Risk: Some Evidence from Density Forecasts, WP04-10
11. Ilias Tsiakas, Periodic Stochastic Volatility and Fat Tails, WP04-09
12. Ilias Tsiakas, Is Seasonal Heteroscedasticity Real? An International Perspective, WP04-08
13. Damin Challet, Andrea De Martino, Matteo Marsili and Isaac Castillo, Minority games with finite score memory, WP04-07
14. Basel Awartani, Valentina Corradi and Walter Distaso, Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average, WP04-06
15. Andrew Patton and Allan Timmermann, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, WP04-05
16. Andrew Patton, Modelling Asymmetric Exchange Rate Dependence, WP04-04
17. Alessio Sancetta, Decoupling and Convergence to Independence with Applications to Functional Limit Theorems, WP04-03
18. Alessio Sancetta, Copula Based Monte Carlo Integration in Financial Problems, WP04-02
19. Abhay Abhayankar, Lucio Sarno and Giorgio Valente, Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability, WP04-01

2002

1. Paolo Zaffaroni, Gaussian inference on Certain Long-Range Dependent Volatility Models, WP02-12
2. Paolo Zaffaroni, Aggregation and Memory of Models of Changing Volatility, WP02-11
3. Jerry Coakley, Ana-Maria Fuertes and Andrew Wood, Reinterpreting the Real Exchange Rate - Yield Diffential Nexus, WP02-10
4. Gordon Gemmill and Dylan Thomas , Noise Training, Costly Arbitrage and Asset Prices: evidence from closed-end funds, WP02-09
5. Gordon Gemmill, Testing Merton's Model for Credit Spreads on Zero-Coupon Bonds, WP02-08
6. George Christodoulakis and Steve Satchell, On th Evolution of Global Style Factors in the MSCI Universe of Assets, WP02-07
7. George Christodoulakis, Sharp Style Analysis in the MSCI Sector Portfolios: A Monte Caro Integration Approach, WP02-06

8. George Christodoulakis, Generating Composite Volatility Forecasts with Random Factor Betas, WP02-05
9. Claudia Riveiro and Nick Webber, Valuing Path Dependent Options in the Variance-Gamma Model by Monte Carlo with a Gamma Bridge, WP02-04
10. Christian Pedersen and Soosung Hwang, On Empirical Risk Measurement with Asymmetric Returns Data, WP02-03
11. Roy Batchelor and Ismail Orgakcioglu, Event-related GARCH: the impact of stock dividends in Turkey, WP02-02
12. George Albanis and Roy Batchelor, Combining Heterogeneous Classifiers for Stock Selection, WP02-01

2001

1. Soosung Hwang and Steve Satchell , GARCH Model with Cross-sectional Volatility; GARCHX Models, WP01-16
2. Soosung Hwang and Steve Satchell, Tracking Error: Ex-Ante versus Ex-Post Measures, WP01-15
3. Soosung Hwang and Steve Satchell, The Asset Allocation Decision in a Loss Aversion World, WP01-14
4. Soosung Hwang and Mark Salmon, An Analysis of Performance Measures Using Copulae, WP01-13
5. Soosung Hwang and Mark Salmon, A New Measure of Herding and Empirical Evidence, WP01-12
6. Richard Lewin and Steve Satchell, The Derivation of New Model of Equity Duration, WP01-11
7. Massimiliano Marcellino and Mark Salmon, Robust Decision Theory and the Lucas Critique, WP01-10
8. Jerry Coakley, Ana-Maria Fuertes and Maria-Teresa Perez, Numerical Issues in Threshold Autoregressive Modelling of Time Series, WP01-09
9. Jerry Coakley, Ana-Maria Fuertes and Ron Smith, Small Sample Properties of Panel Time-series Estimators with I(1) Errors, WP01-08
10. Jerry Coakley and Ana-Maria Fuertes, The Felsdtein-Horioka Puzzle is Not as Bad as You Think, WP01-07
11. Jerry Coakley and Ana-Maria Fuertes, Rethinking the Forward Premium Puzzle in a Non-linear Framework, WP01-06
12. George Christodoulakis, Co-Volatility and Correlation Clustering: A Multivariate Correlated ARCH Framework, WP01-05
13. Frank Critchley, Paul Marriott and Mark Salmon, On Preferred Point Geometry in Statistics, WP01-04
14. Eric Bouyé and Nicolas Gaussel and Mark Salmon, Investigating Dynamic Dependence Using Copulae, WP01-03
15. Eric Bouyé, Multivariate Extremes at Work for Portfolio Risk Measurement, WP01-02
16. Erick Bouyé, Vado Durrleman, Ashkan Nikeghbali, Gael Riboulet and Thierry Roncalli, Copulas: an Open Field for Risk Management, WP01-01

2000

1. Soosung Hwang and Steve Satchell , Valuing Information Using Utility Functions, WP00-06
2. Soosung Hwang, Properties of Cross-sectional Volatility, WP00-05
3. Soosung Hwang and Steve Satchell, Calculating the Miss-specification in Beta from Using a Proxy for the Market Portfolio, WP00-04
4. Laun Middleton and Stephen Satchell, Deriving the APT when the Number of Factors is Unknown, WP00-03
5. George A. Christodoulakis and Steve Satchell, Evolving Systems of Financial Returns: Auto-Regressive Conditional Beta, WP00-02
6. Christian S. Pedersen and Stephen Satchell, Evaluating the Performance of Nearest Neighbour Algorithms when Forecasting US Industry Returns, WP00-01

1999

1. Yin-Wong Cheung, Menzie Chinn and Ian Marsh, How do UK-Based Foreign Exchange Dealers Think Their Market Operates?, WP99-21
2. Soosung Hwang, John Knight and Stephen Satchell, Forecasting Volatility using LINEX Loss Functions, WP99-20
3. Soosung Hwang and Steve Satchell, Improved Testing for the Efficiency of Asset Pricing Theories in Linear Factor Models, WP99-19
4. Soosung Hwang and Stephen Satchell, The Disappearance of Style in the US Equity Market, WP99-18
5. Soosung Hwang and Stephen Satchell, Modelling Emerging Market Risk Premia Using Higher Moments, WP99-17
6. Soosung Hwang and Stephen Satchell, Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets, WP99-16
7. Soosung Hwang, The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and their Finite Sample Properties, WP99-15
8. Ronald MacDonald and Ian Marsh, Currency Spillovers and Tri-Polarity: a Simultaneous Model of the US Dollar, German Mark and Japanese Yen, WP99-14
9. Robert Hillman, Forecasting Inflation with a Non-linear Output Gap Model, WP99-13
10. Robert Hillman and Mark Salmon , From Market Micro-structure to Macro Fundamentals: is there Predictability in the Dollar-Deutsche Mark Exchange Rate?, WP99-12
11. Renzo Avesani, Giampiero Gallo and Mark Salmon, On the Evolution of Credibility and Flexible Exchange Rate Target Zones, WP99-11
12. Paul Marriott and Mark Salmon, An Introduction to Differential Geometry in Econometrics, WP99-10
13. Mark Dixon, Anthony Ledford and Paul Marriott, Finite Sample Inference for Extreme Value Distributions, WP99-09
14. Ian Marsh and David Power, A Panel-Based Investigation into the Relationship Between Stock Prices and Dividends, WP99-08
15. Ian Marsh, An Analysis of the Performance of European Foreign Exchange Forecasters, WP99-07
16. Frank Critchley, Paul Marriott and Mark Salmon, An Elementary Account of Amari's Expected Geometry, WP99-06
17. Demos Tambakis and Anne-Sophie Van Royen, Bootstrap Predictability of Daily Exchange Rates in ARMA Models, WP99-05
18. Christopher Neely and Paul Weller, Technical Analysis and Central Bank Intervention, WP99-04
19. Christopher Neely and Paul Weller, Predictability in International Asset Returns: A Re-examination, WP99-03
20. Christopher Neely and Paul Weller, Intraday Technical Trading in the Foreign Exchange Market, WP99-02
21. Anthony Hall, Soosung Hwang and Stephen Satchell, Using Bayesian Variable Selection Methods to Choose Style Factors in Global Stock Return Models, WP99-01

1998

1. Soosung Hwang and Stephen Satchell, Implied Volatility Forecasting: A Compaison of Different Procedures Including Fractionally Integrated Models with Applications to UK Equity Options, WP98-05
2. Roy Batchelor and David Peel, Rationality Testing under Asymmetric Loss, WP98-04
3. Roy Batchelor, Forecasting T-Bill Yields: Accuracy versus Profitability, WP98-03
4. Adam Kurpiel and Thierry Roncalli , Option Hedging with Stochastic Volatility, WP98-02
5. Adam Kurpiel and Thierry Roncalli, Hopscotch Methods for Two State Financial Models, WP98-01