Modelling 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

This draft: November 2001

Abstract

A large empirical literature has reported that the futures market contains valuable information for explaining stock returns and that stock returns display significant cross-correlations internationally. A parallel literature has recorded evidence that the distribution of stock returns is close to a mixture of normal distributions and that Markov switching models may therefore provide an adequate characterization of stock returns data. This paper ties together these strands of research in that we propose a vector equilibrium correction model of stock returns that exploits the information in the futures market, while also allowing for regime-switching behavior and international spillovers across stock market indices. Using data for three major stock market indices since 1988, we find that our model significantly outperforms a number of alternative models in sample on the basis of standard statistical criteria. In an out-of-sample forecasting exercise, the model produces some of the highest R^2 hitherto recorded in the literature and beats all of the competing models considered on the basis of density forecast accuracy.

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, whose hospitality is gratefully acknowledged. We are grateful to Abhay Abhyankar and Lutz Kilian for useful conversations and comments on previous drafts. The views expressed in the paper should not be seen as representative of the position of the Federal Reserve Bank of St. Louis or the Board of Governors of the Federal Reserve System.

1 Introduction

A large body of both theoretical and empirical research focusing on modelling 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).[1] 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.[2] This literature, discussed in greater detail in the next section, has generally reported evidence that the futures market contains valuable information for modelling 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).[3]

Alongside the work on modelling 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). Also, not only Markov-switching models fit stock returns data well, but they have often 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).[4]

In this paper, we tie together pieces of these somewhat disparate albeit related strands of

[1] Related work as employed structural vector autoregressive models; see, for example, Lee (1998).

[2] 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.

[3] 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. See also Lee (1995).

[4] Other studies in this literature have provided ample empirical evidence that the dynamic relationship linking stock and futures prices may be characterized by significant nonlinearities that can be well char-

research. 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 ¯t and out-of-sample forecasting, over conventional 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 prices and futures prices that is based on an extension of Markovian regime shifts to a nonstationary framework, for which the underlying econometric theory has recently been developed. Given the evidence of signi¯cant 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 modelling and forecasting, even though this involves estimating and forecasting from a sophisticated multivariate nonlinear model.

Using weekly data since 1988 for three major stock market indices - the S&P 500, the NIKKEI 225 and the FTSE 100 indices - we con¯rm that the futures market does contain valuable information to explain stock returns in a linear VECM framework. However, we show that conventional linear VECMs, even when allowing for international spillovers in the equilibrium correction equations, display signi¯cant residual nonlinearity and are strongly rejected when tested against the alternative of an MS-VECM. Thus, we show that allowing for nonlinearities and for international spillovers in an MS-VECM results in a superior empirical model which explains a large proportion of the stock returns examined over the sample. Finally, we 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 as well as on the ability of the models to forecast the true predictive density of stock returns out of sample.[5] 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; Timmerman, 2000). To anticipate our forecasting results, we ¯nd that the MS-VECM that allows for international spillovers does not outperform the competing models examined in terms of point forecasting performance, even though it generates a remarkably high R^2 out of sample. However, our model signi¯cantly outperforms all of the competing models in terms of density forecasting performance in that it generates predictive densities that are much closer to the true predictive density of the data.

The remainder of the paper is set out as follows. In Section 2 we describe our empirical framework for modelling stock and futures prices allowing for international spillovers and nonlinear dynamics. We also brie°y set out the econometrics of Markov-switching multivariate models as applied to nonstationary processes and cointegrated systems. In Section 3 we report our empirical testing and estimation results, while in the subsequent section we report our forecasting results. A ¯nal section concludes.

---

acterized using threshold models of various sort. These nonlinearities are rationalized on the basis of a number of di®erent factors such as non-zero transactions 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; Gao and Wang, 1999).

[5]By true predictive density of the data we mean the density of the data estimated 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 is not estimated over the full sample but only over the forecast period.

2 Modelling stock returns: an empirical framework

In this section we outline our empirical framework for modelling stock returns, which we apply to our data in the subsequent sections. First, we use a simple variant of the 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 system of VECMs which explicitly allows for both regime shifts and international spillovers in 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. Consider, for example, a market containing an asset, a stock index, whose price $S(t)$ under the equivalent martingale measure evolves according to:

$$dS(t) = S(t)(r - q)dt + \sigma_S dW_S(t); \tag{1}$$

where r is the (constant) risk-free interest rate; q is the (constant) dividend yield on the index; σ_S is the volatility of the index; $W_S(t)$ is a one-dimensional standard Brownian motion in a complete probability space.

Standard derivatives pricing theory gives the following expression for the futures price $F(t;T)$ at time t for delivery of the stock at time $T \geq t$:

$$F(t;T) = E[S(T)|\mathcal{F}(t)]; \tag{2}$$

where E denotes the mathematical expectation with respect to the martingale measure P, and $\mathcal{F}(t)$ denotes the information set at time t (e.g. see Karatzas and Shreve, 1998). Given (1)-(2), the futures price has the well-known formula:

$$F(t;T) = S(t)\exp(r_c(T - t)), \tag{3}$$

where $r_c = r - q$. This is the familiar expression for a futures price in a non-random interest rate environment.[6]

Taking logarithms of both sides of equation (3) and rearranging yields:

$$[\log F(t;T) - r_c(T - t)] - \log S(t) = 0. \tag{4}$$

If the logarithm of the futures price adjusted for interests and dividends and the logarithm of the stock price are both unit root or I(1) processes, then equation (4) implies that the adjusted log-futures price and the log-stock price move together. In turn, this implies

[6] Alternatively, equation (3) can be derived without assuming the process for $S(t)$ given in (1), by assuming that the futures price implied by the cost of carry model is $F(t;T) = S(t)\,[\exp r_c(T - t)]\,(1 - \delta(T - t))$. If $\delta(T - t)$ is small, then $(1 - \delta(T - t)) \approx \exp -\delta(T - t)$ and the equation for the futures price can be approximated by equation (3) (e.g. see Dwyer, Locke and Yu, 1996).

that adjusted futures and stock prices exhibit a common stochastic trend and are cointegrated with cointegrating vector $[1; -1]$. In our empirical work, we subtract the weekly mean from the logarithms of the futures and stock prices. Demeaning the futures price removes the constant part of the dividends and interest rates for that week, so that the demeaned logarithms of the futures and stock prices should cointegrate with cointegrating vector $[1; -1]$. Also, the difference between the demeaned logarithms of the futures and stock prices is the deviation of the futures basis from its weekly mean.[7]

Further, by the Granger Representation Theorem (Engle and Granger, 1987) the futures and stock prices must possess a VECM representation where the adjusted (i.e. demeaned) futures basis plays the part of the equilibrium error. We exploit this framework and use exactly a VECM representation to demonstrate that a large amount of information may be extracted from the futures market in order to forecast stock returns.

2.2 Regime-switching equilibrium correction in stock index futures markets

A large literature has documented convincing 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 Åsbrink, 1998; Susmel, 1999). In fact, the relevant literature suggests that not only Markov-switching models fit stock price data well, but they often perform very 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 superior stock returns forecasts relative to several alternative specifications. This is done through estimating a fairly general MS-VECM for stock prices and futures prices which is essentially 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:

$$y_t = \nu(z_t) + \sum_{i=1}^{p} \Pi_i y_{t-i} + \varepsilon_t; \tag{5}$$

where y_t is a K-dimensional observed time series vector, $y_t = [y_{1t}; y_{2t}; \ldots; y_{Kt}]'$; $\nu(z_t) = [\nu_1(z_t); \nu_2(z_t); \ldots; \nu_K(z_t)]'$ 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, $\varepsilon_t \sim NID(0; \Sigma)$. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of

[7] The logarithmic basis $b(t; T)$ at time t is defined as $\log B(t; T) = \log F(t; T) - \log S(t)$.

states $z_t \in \{1, \ldots, M\}$ governed by the transition probabilities $p_{ij} = \Pr(z_{t+1} = j \mid z_t = i)$, and $\sum_{j=1}^{M} p_{ij} = 1 \ \forall i, j \in \{1, \ldots, M\}$.[8]

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 the 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 (5) implies an MS-VECM of the form:

$$\Delta y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t; \tag{6}$$

where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ are matrices of parameters, and $\Pi = \sum_{i=1}^{p} \Pi_i - I$ is the long-run impact matrix whose rank r determines the number of cointegrating vectors (e.g. Johansen, 1995; Krolzig, 1999).[9]

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 futures markets analyzed, the futures price and the stock price (hence $y_t = [f_t, s_t]'$) where f_t and s_t denote the demeaned logarithmic futures and stock prices respectively. Following the reasoning of our discussion in Section 2.1, a unique cointegrating relationship should exist between f_t and s_t. 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 as well as in the variance-covariance matrix. This model, the Markov-Switching-Intercept-Heteroskedastic-VECM or MSIH-VECM, may be written as follows:

$$\Delta y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t; \tag{7}$$

where $\Pi = \alpha\beta'$, $u_t \sim NIID(0, \Sigma(z_t))$ and $z_t \in \{1, \ldots, M\}$.

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 modelling 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

[8] To be precise, z_t is assumed to follow an ergodic M-state Markov process with transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix},$$

where $p_{iM} = 1 - p_{i1} - \cdots - p_{i,M-1}$ for $i \in \{1, \ldots, M\}$.

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

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 modelling stock returns

Although conventional time series models employed to explain or forecast stock returns treat a particular asset or index in isolation, a vast empirical and theoretical 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 (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.

Table 1, which reports correlation coefficients for the three different futures bases examined in this study (the S&P 500, the FTSE 100, and the NIKKEI 225), provides clear evidence in favor of the conjecture that the three futures bases of these indices display substantial and statistically significant cross-correlation, ranging from about 21 percent for the S&P 500-NIKKEI 225 pair to about 58 percent for the NIKKEI 225-FTSE 100 pair. These correlations indicate clear interdependencies between these three major stock market indices, suggesting that the allowance for spillovers in our spot-futures VECM may yield substantial improvements relative to individual VECM estimation due to the incremental information yielded by the cross-correlation of the indices examined.

This line of reasoning suggests the possibility of enriching our MS-VECM framework by allowing for spillovers through the equilibrium correction terms, that is by allowing for 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), Granger and Haldrup (1997) - which therefore provides a useful way of describing formally the above ideas.

Consider the MS-VECM (6) 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 (6) 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:

$$\Delta Y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \alpha\beta' Y_{t-1} + \varepsilon_t; \qquad (8)$$

where Γ_i is an $n \times n$ matrix of autoregressive parameters, α and β' 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 β' can be factorized as

$$\beta' = \begin{bmatrix} \beta'_{11} & 0 & 0 \\ 0 & \beta'_{22} & 0 \\ 0 & 0 & \beta'_{33} \end{bmatrix} \tag{9}$$

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

$$\alpha = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix} \tag{10}$$

we have type B-separation or separation in the equilibrium correction, whereas if we factorize the matrix Γ_i as

$$\Gamma_i = \begin{bmatrix} \Gamma^j_{11} & 0 & 0 \\ 0 & \Gamma^j_{22} & 0 \\ 0 & 0 & \Gamma^j_{33} \end{bmatrix} \tag{11}$$

we have type A-separation or separation in the dynamic adjustment towards the long-run equilibrium defined for each y^j for $j = 1; 2; 3$ (e.g. Granger and Haldrup, 1997). If all of the conditions (9)-(11) hold there is complete separation, while if condition (9) is associated with either (10) or (11) we have partial separation.

Our earlier discussion on spillovers in the dynamics of stock market returns is consistent with a situation where, although two or more different stock indices are `separated in the long-run' (i.e. condition (9) holds), there may be important short-run relationships between them and, therefore, the deviation from the equilibrium relationship (defined by the futures basis) from one index may enter the equilibrium correction equation of another index (i.e. condition (11) 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 the stock and futures prices for index k, as predicted by the standard cost of carry model. Hence, stock and futures prices for any other index $j \neq k$ do not enter the long-run cointegration 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, where we start by estimating cointegrating relationships and, therefore, equilibrium correction terms, which imply plausible parameters and are consistent with the definition of the futures basis (i.e. with the $[1; -1]$ cointegrating vector implied by the framework in Section 2.1). Thus, 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: modelling

3.1 Data, preliminary statistics and cointegration analysis

The data set comprises weekly time series on futures 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 were obtained from Datastream, and the sample period examined spans from September 1988 to December 2000. We use this sample period because the NIKKEI 225 stock index futures was ¯rst traded on September 3 1988 in the Osaka Stock Exchange (OSE).[10] In our empirical work, we carried out estimations over the period September 1988-December 1998, reserving the last two years of data for out-of-sample forecasting tests.

Panel A of Table 2 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 ¯rst moment of the futures price is larger than the ¯rst 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 2 up to order 12, suggest that each spot and futures price examined displays very strong ¯rst-order serial correlation, while none of these series appears to be signi¯cantly serially correlated at higher lags.[11]

As a preliminary exercise, we tested for unit root behavior of each of the (log) futures price and spot price time series by calculating standard augmented Dickey-Fuller test statistics.[12] In each case, the number of lags was chosen such that no residual autocorrelation was evident in the auxiliary regressions. In keeping with the very large number of studies of unit root behavior for these time series and conventional ¯nance theory, we were in each case unable to reject the unit root null hypothesis at conventional nominal levels of signi¯cance. On the other hand, di®erencing 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.

The implementation of the Johansen (1988, 1991) maximum likelihood cointegration procedure is essentially the ¯rst stage of the two-stage procedure designed to estimate an

[10] More precisely, NIKKEI 225 futures contracts were ¯rst 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).

[11] Considering that some of the papers in a related literature were engendered by considerations of microctructural nature and focused on price and basis changes (e.g. Miller, Muthuswamy and Whaley, 1994), we also examined the autocorrelation function (up to 12 lags) of spot price changes, futures price changes, and basis changes for all indices studied. The results (not reported to conserve space but available upon request) indicated that changes in both spot and futures prices for each index display signi¯cant autocorrelation, fairly strong especially at lag 1. Further, for each index, changes in the futures basis display the well-documented negative autocorrelation. As suggested by Miller, Muthuswamy and Whaley (1994), the negative autocorrelation of the basis has important implications for microstructural ¯nance, and - at least at high frequency - may be generated by the fact that index stocks do not trade continuously.

[12] The results are not reported to save space, but they are available from the authors upon request.

MS-VECM for Δf_t and Δs_t, as discussed in Section 2.2. We employed the Johansen procedure in a VAR allowing for a maximum lag length of five and an unrestricted constant term, hence testing for cointegration in the long-run model:[13]

$$f_t - \phi s_t = \vartheta. \tag{12}$$

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 restriction suggested by framework described in Section 2.1 that the cointegrating parameter associated with s_t equals unity ($\phi = 1$) could not be rejected at conventional nominal levels of significance for each of the estimated VARs.[14] Given the restriction $\phi = 1$, the cointegrating residuals may be seen as the estimated deviation of the basis from its mean.

Overall, the results in this section suggest that a unique cointegrating relationship exists between f_t and s_t for each of the three major stock indices examined, lending support to the simple framework outlined in Section 2.1.

3.2 Linear dynamic modelling and linearity tests

Preliminary to considering an MS-VECM, we estimated a standard linear bivariate VECM for Δf_t and Δs_t, which is implied by the finding of cointegration between f_t and s_t reported in the previous sub-section (Granger, 1986; Engle and Granger, 1987). Thus, using full-information maximum likelihood (FIML) methods, we estimated for each stock index a bivariate VECM of the form

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t \tag{13}$$

where $y_t = [f_t; s_t]'$; we allowed for a maximum lag length of five as suggested by both the Akaike information criterion (AIC) and the Schwartz information criterion (SIC). Employing the conventional general-to-specific procedure, we obtained, for each stock index examined, fairly parsimonious models for Δf_t and Δs_t which display no residual serial correlation.[15]

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

$$\Delta Y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-1} + \varepsilon_t \tag{14}$$

[13] The VAR considered is essentially model $H_1^*(r)$ in Johansen (1995, p. 83) notation, where a linear deterministic trend is implicitly allowed for but this can be eliminated by the cointegrating relations and the process contains no trend-stationary components; hence, the model allows for a linear trend in each variable but not in the cointegrating relations.

[14] LR tests of the hypothesis $\phi = 1$ could not be rejected with p-values equal to 0.616, 0.587 and 0.603 for the S&P 500, the NIKKEI 225 and the FTSE 100 respectively.

[15] Full details on these estimation results are available from the authors upon request, but are not reported to conserve space and because the main focus of this section of the paper is on linearity testing applied to the linear VECM residuals.

where $Y_t = [f_t^{SP500}, s_t^{SP500}, f_t^{NK225}, s_t^{NK225}, f_t^{FTSE100}, s_t^{FTSE100}]'$. We tested for type B-separation (separation in the equilibrium correction) by estimating model (14) imposing the zero restrictions as in (10) and executing a standard likelihood ratio (LR) test. The results allow us to reject the zero restrictions under the null hypothesis (10), 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 (13) and (14), namely Ramsey's (1969) RESET test and the Brock, Dechert and Sheinkman (BDS) (1991) test for testing the null hypothesis that the residuals from (13) and (14) are independent and identically distributed (iid) against an unspecified alternative.[16] Application of both of these tests provided strong empirical evidence that the linear VECM fails to capture important nonlinearities in the data generating process, as linearity is rejected with marginal significance levels (p-values) of virtually zero (see Panels a) and b) of Table 4).[17]

3.3 MS-VECM estimation results

Next, we applied the conventional "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.[18] However, for any MS-VECM estimated the implicit assumption that the regime shifts affect only the drift term of the VECM was found to be inappropriate. In fact, we checked the relevance of conditional heteroskedasticity by estimating an MS-VECM where the Gaussian innovation is allowed to be heteroskedastic, $\varepsilon_t \sim NID(0, \Sigma(z_t))$. We then tested the hypothesis of no regime dependence in the variance-covariance matrix using an LR test of the type suggested by Krolzig (1997, p. 136). The results suggest very strong rejection of the null of no regime dependence, clearly indicating that an MS-VECM that allows for shifts in both the intercept v and

[16]Under the RESET test statistic, the alternative model involves a higher-order polynomial to represent a different functional form; under the null hypothesis, the statistic is distributed as $\chi^2(q)$ with q equal to the number of higher-order terms in the alternative model. The BDS test for a series u_t is calculated in the following way. Let $u_{t,v}$ be a set of consecutive terms from u_t : $u_{t,v}\ \{u_t, u_{t+1}, \ldots, u_{t+v-1}\}$. The pair of vectors $u_{t,v}$ and $u_{s,v}$ are said to be no more than ς apart if $|u_{t+j} - u_{s+j}| \leq \varsigma$ for $j = 0, 1, \ldots, v-1$. Thus, the correlation integral $C_v(\varsigma)$ is defined as the product of the limit of T^{-2} (T being the number of observations) times the number of ς-close pairs (s, t), essentially measuring the probability that the pairs of points (s, t) are within ς of each other. The BDS statistic is then constructed as $S(v, \varsigma) = \hat{C}_v(\varsigma) - [\hat{C}_1(\varsigma)]^v$ for some v and ς. Under the null hypothesis that u_t is iid, $\sqrt{T}[S(v, \varsigma)] \sim N(0, \xi)$, where the variance ξ is a function of v and ς. Rejection of the null implies that some form of nonlinearity is present in u_t, although the type of nonlinearity cannot be exactly determined under the BDS test. BDS (1991) suggest that the choice of v and, particularly, the choice of ς, are crucial for the power of the test, which is reasonably powerful only in large samples. BDS (1991) also suggest values of ς between 0.5 and 1.5 times the standard deviation of u_t, whereas the value of v should preferably be such that $(T/v) > 200$.

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

[18]Essentially the bottom-up procedure consists of starting with a simple but statistically reliable Markov switching model by restricting the effects of regime shifts on a limited number of parameters and check the model against alternatives. In such a procedure, most of the structure contained in the data is not attributed to regime shifts, but explained by observable variables, therefore being consistent with the general-to-specific approach to econometric modelling. For a comprehensive discussion of the bottom-up procedure, see Krolzig (1997).

the variance-covariance matrix Σ, that is an MSIH(M)-VECM(p), is the most appropriate model within its class in the present application. Then, we 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.[19] 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; Richmond and Susmel, 1998a,b), the presence of two regimes for each stock index.

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

$$\begin{aligned} \Delta y_t &= \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim NID(0; \Sigma(z_t)) \qquad z_t = 1; 2 \end{aligned} \tag{15}$$

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

$$\begin{aligned} \Delta Y_t &= \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim NID(0; \Sigma(z_t)) \qquad z_t = 1; 2. \end{aligned} \tag{16}$$

Making no assumption on the relationship between the regime shifts occurring in the stock indices examined (see Krolzig, 1997; Hamilton and Lin, 1996) implies that the number of regimes incorporated in model (16), and consequently the dimension of the transition matrix, is $2^3 = 8$.[20]

The empirical results are very encouraging on a number of fronts. The estimation yields plausible results for each VECM estimated, with all coefficients found to be strongly statistically significantly different from zero at conventional levels of significance.[21] The impact effect of regime shifts also appears to be substantial on the variance-covariance matrix $\Sigma(z_t)$. Furthermore, we computed an LR test statistic for linearity, which essentially tests the hypothesis that the true model is a linear VECM against the alternative of the MSIH-VECM reported in Table 5. 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 that allows for shifts in ν and Σ in the econometric modelling of the data under examination.[22] Moreover, even in the context of Markov-switching models, type B-separation is rejected by the data. In

[19] One may argue that the lag length of unity which was selected by the bottom-up procedure may be too low, given that we allowed for a maximum lag length of five in estimating the linear VECM. Nevertheless, estimation of an MSIH-VECM with higher-order lags provided several insignificant autoregressive parameters and did not change qualitatively any of the results reported below. Further, given that the diagnostic tests and the graphs of the residuals from the MSIH(3)-VECM(1) did not indicate either misspecification or residual serial correlation and that the coefficients of determination were found to be very high, we decided to stick to the more parsimonious models.

[20] For further technical details see Appendix A.

[21] The full set of estimated coefficients are not reported here to save space, but full details are available upon request.

[22] Note 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

fact the likelihood ratio test (LR2) reported in the second column of Table 5 strongly rejects the null of separation in the equilibrium correction terms.

We also compute coe±cients of determination (R^2 and $\overline{R}^2$); the $\overline{R}^2$ was 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-dependent heteroskedasticity. The results are reported in Panel a) of Table 6. Under this measure of goodness of ¯t, two facts arise. First, the role of non-separation in the equilibrium correction terms is important to explain the variability of futures and spot returns: columns 2 and 4 highlight the improvement in the in-sample predictive performance of the models when the futures bases from di®erent stock markets are incorporated as explanatory variables in the returns equations. Second, the role of nonlinearities appears to be very important to better explain stock returns. Columns 3 and 4 show how nonlinearities of the type speci¯ed in Section 2 help to capture the general features exhibited by the time series under investigation. Thus, examining the last column of Panel a) of Table 6, where international spillovers and nonlinearities are both explicitly taken into account, suggests that the in-sample performance of the model is highly satisfactory. The uncorrected R^2 exhibits values larger than 45% for all futures and spot indices. Even correcting for the large number of parameters of the MSIH(8)-VECM(1) model, the coe±cient of determination is at least twice as large as the coe±cient of determination obtained for the bivariate MS-VECM models and more than 10 times larger than the coe±cient of determination of the standard linear VECM models.[23] [24]

4 Empirical analysis II: forecasting

4.1 R^2 out of sample and point forecasting performance

One of our results, corroborating some previous ¯ndings 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 usefulness of our nonlinear MS-VECM characterization and the gain from using a sophisticated nonlinear empirical model, dynamic out-of-sample forecasts of stock returns were constructed using the MSIH-VECM estimated and discussed in the previous section. In particular, we calculated one-step-ahead forecasts over the period January 1999-December 2000.[25] The out-of-sample forecasts for a given horizon are constructed according to a recursive procedure that is conditional only upon information available up to the date of the

valid if the null model is a linear model with iid errors; in the present case, it is di±cult 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 di®er from the adjusted Â^2 distribution proposed by Davies (1977, 1987). For extensive discussions of the problems related to LR testing in this context, see Hansen (1992, 1996) and Garcia (1998). We are thankful to Bruce Hansen for clarifying several econometric issues related to LR testing in the present context.

[23]Note that all of our estimated MS-VECMs are clearly stationary, as con¯rmed by calculating the value of the spectral radius as in Karlsen (1990).

[24]As a way to evaluate the dynamic properties of the estimated Markov-switching models we also examined the e®ects 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.

[25]For a description of the econometric issues related to out-of-sample forecasting in a Markov-switching framework, see Hamilton (1993) and Krolzig (2000).

forecasts and with successive re-estimation as the date on which forecasts are conditioned moves through the data set.

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 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 (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 different criteria. Panel a) of Table 7 shows the R^2 out-of-sample, the mean absolute errors (MAE) and the mean square errors (MSE) for each of the estimated models. The pattern described by the R^2 is encouraging. The MSIH-VECM (16) (i.e. the nonlinear VECM which allows for international spillovers) exhibits the highest goodness of fit out-of-sample for each of the indices examined: the R^2 out-of-sample is always higher than 10 percent and is close to 20% in the case of the NIKKEI 225 index. Also, the difference between the R^2 out-of-sample from the MSIH-VECM and the R^2 out-of-sample obtained from each of the alternative models is large, suggesting that both nonlinearities and spillovers are important to explain, even out-of-sample, the dynamics of stock returns.

This scenario is not confirmed, however, by the analysis of the MAEs and MSEs. Indeed, on the basis of these criteria, the MSIH-VECM (16) is slightly better (produces lower MAEs and MSEs) only for the S&P 500 index, while for the other two stock indices examined it performs slightly worse than the alternative models. However, the results of the Diebold and Mariano (1995) test, reported in the Panel b) of Table 7, 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 MSEs reported in Table 7 are not statistically significant.[26] [27]

Overall, the results from analyzing the R^2 out of sample, which favor the MSIH-VECM (16), appear to conflict with the results from analyzing MAEs and MSEs, which do not support any particular model.

4.2 Density forecasting performance

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

[26] A consistent estimate of the spectral density at frequency zero $\hat{f}(0)$ 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 $\hat{\rho}$ and the innovation variance from the AR(1) process $\hat{\sigma}^2$. Then the optimal truncation lag A for the Parzen window in the Newey and West estimator is given by the Andrews' rule $A = 2.6614\left[\hat{\alpha}(2)T\right]^{1/5}$ where $\hat{\alpha}(2)$ is a function of $\hat{\rho}$ and $\hat{\sigma}^2$. The Parzen window has been chosen according to Gallant (1987, p. 534).

[27] 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). Also, note that the non-rejection of the null of equal point forecast accuracy may be due to the well documented low power of the Diebold-Mariano test statistic in finite sample (see Kilian and Taylor, 2001, and the references therein).

criteria to evaluate their predictive accuracy we obtained somewhat conflicting results. How can one reconcile, for example, the finding of an R^2 out-of-sample larger than 12% or so and average improvements ranging between 5 and 12 times relative to the alternative models with our inability to beat the alternative models on the basis of MAEs and MSEs?

One possible explanation of these findings is suggested by careful examination of Table 8, which reports the absolute differences between the first four moments of the estimated forecast densities and the true predictive density of the data. The figures in Table 8 suggest that focusing only on the first two moments of the stock returns distributions, which is effectively what one does when using point forecast accuracy tests, we do not exploit the whole information provided by the MS-VECMs out-of-sample predictions. In particular, the MSIH-VECM (16) appears to 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. This evidence is made visually clear by Figures 1-3, which plot, for each of the stock indices examined, the forecast densities implied by each of the competing models together with the true predictive density of stock returns. The graphs make simply too apparent how the MSIH-VECM (16) produces density forecasts much closer to the true predictive density of the data than any of the other competing models. In some sense, therefore, inspection of the absolute differences given in Table 8 and the forecast densities plotted in Figures 1-3 may help us to reconcile the puzzling evidence of the high R^2 out-of-sample and the unsatisfactory results in terms of MAEs and MSEs documented in the previous subsection.

A logical next step then involves testing formally the hypothesis that the forecast density implied by the MSIH-VECM (16) is the closest to the true predictive density of the data in order to add econometric precision to the informal evidence provided by Table 8 and Figures 1-3. 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; Timmerman, 2000; Pesaran and Skouras, 2001). 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, this line of research has produced several methods either to measure the closeness of two density functions or to test the hypothesis that the predictive density generated by a particular model corresponds to the true predictive density. However, these tests do not allow us to test the null hypothesis of equal density forecast accuracy between competing models.

In order to test formally the null hypothesis of equal density forecast accuracy between the MS-VECM (16) and each of the alternative models considered we employed the η-test recently proposed by Sarno and Valente (2001) for evaluating the accuracy of the density forecasts generated by competing models. The η-test is similar in spirit to the test suggested by Diebold and Mariano (1995) but involves the analysis of the whole forecast density instead of point forecasts.[28] The η-test statistic is constructed as follows:

[28] See Appendix B for details on the calculation and the properties of the η-test.

$$\eta = \frac{\bar{d}}{\sqrt{\frac{\hat{\sigma}}{B}}} \tag{17}$$

where $\bar{d} = \frac{1}{B}\sum_{j=1}^{B} d^j = \frac{1}{B}\sum_{j=1}^{B}\left(\widehat{ISD}_1^j - \widehat{ISD}_2^j\right)$ is an average (over $j = 1, \ldots, B$ bootstrap replications) of the difference between two estimated integrated square differences, $\widehat{ISD}_1^j$ and $\widehat{ISD}_2^j$; $\widehat{ISD}_1^j$ refers to the integrated square difference between the MSIH-VECM (16) and the true predictive density, whereas $\widehat{ISD}_2^j$ refers to the integrated square difference between the particular competing model considered and the true predictive density; $\hat{\sigma}$ is a consistent bootstrap estimate of the variance of the difference d.[29] Under general conditions, the η-test statistic is distributed as standard normal under the null hypothesis that the two competing models have equal density forecast accuracy.

The results from applying the η-test to our models, reported in Table 9, confirm the informal evidence documented in Table 8 and Figures 1-3 since the null of equal density forecast accuracy is strongly rejected in each case. More precisely, for each alternative model and each stock index examined, the MSIH-VECM (16) produces the best density forecasts, with the null hypothesis of equal density forecast accuracy being rejected with p-values of virtually zero.

Summing up, the forecasting results in this section suggest that the general MSIH-VECM that allows for international spillovers performs significantly better than any other linear and nonlinear model considered in this paper in terms of explaining the out of sample behavior of stock returns. In particular, while the forecasting performance of the general MSIH-VECM is not statistically different from the performance of the alternative models in terms of point forecasting, the MSIH-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.

5 Conclusion

This article has re-examined the dynamic relationship between spot and futures prices in stock index futures markets using data since 1988 at weekly frequency for three major stock market indices - the S&P 500, the NIKKEI 225 and the FTSE 100 indices. In particular, motivated by several related strands of research, we proposed a nonlinear, Markov-switching vector equilibrium correction model that explicitly takes into account the large amount of evidence that the conditional distribution of stock returns is well characterized by a mixture of normal distributions. Also, we used the recently developed 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 strong evidence in favor of the existence of strong 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

[29] The estimated integrated square difference $\widehat{ISD} = \int\left[\hat{\phi}(x) - \hat{\gamma}(x)\right]^2 dx$ is obtained by estimating the density functions ϕ and γ by means of the Gaussian kernel estimator.

the futures basis and the fundamental predictions of the cost of carry model. Linear vector equilibrium correction models were rejected when tested against a Markov-switching vector equilibrium correction model which allows for shifts in the intercept and the variance-covariance matrix, suggesting the need for a nonlinear, regime-switching speci¯cation. Our preferred nonlinear speci¯cation explains a very large fraction of the stock returns examined, with the R^2 ranging from 0.45 for the FTSE 100 index returns to 0.65 for the S&P 500 index returns, and with the $\overline{R}^2$ ranging from 0.18 for the NIKKEI 225 index returns to 0.26 for the S&P 500 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 future 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 are non-normally distributed and that the nonlinear models employed in this paper imply non-normal predictive densities. In order to adequately measure the forecasting ability of our nonlinear model we employed a test of the null hypothesis of equal density forecast accuracy, which revealed that the forecast density implied by our preferred speci¯cation is, for each stock index considered, signi¯cantly closer to the true predictive density of the data than the forecast densities generated by each of the several competing models.

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 the model used here is very general and °exible, 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. These challenges remain on the agenda for future research.

Table 1. Correlation matrix among futures bases

	S&P 500	NIKKEI 225	FTSE 100
S&P 500	1 [–]		
NIKKEI 225	0.21046 [1.85×10^{-3}]	1 [–]	
FTSE 100	0.39323 [1.95×10^{-3}]	0.58048 [1.15×10^{-3}]	1 [–]

Notes: Estimated total correlation and standard errors (in brackets) are reported. Standard errors of the estimated correlation coefficients are calculated by bootstrap. Data are resampled with replacement 5,000 times.

Table 2. Preliminary data statistics

	S&P 500: f_t	S&P 500: s_t	NIKKEI 225: f_t	NIKKEI 225: s_t	FTSE 100: f_t	FTSE 100: s_t
Minimum	5.5084	5.5081	4.1095	4.1099	7.4325	7.4458
Maximum	6.8977	6.8835	4.5967	4.5901	8.5864	8.5725
Mean	6.0969	6.0921	4.3122	4.3103	7.9467	7.9387
Std Dev	0.1171	0.1172	0.1081	0.1098	0.0764	0.0777
PACF:						
lag 1	0.9915	0.9917	0.9897	0.9899	0.9904	0.9912
lag 2	-0.0134	-0.0146	0.0122	0.0104	-0.0227	-0.0381
lag 3	-0.0388	-0.0368	-0.0246	-0.0130	-0.0182	-0.0202
lag 4	-0.0105	-0.0117	0.0114	-0.0225	0.0158	0.0159
lag 5	0.0373	0.0309	0.0113	0.0139	0.0226	0.0183
lag 6	0.0070	0.0079	-0.0397	-0.0427	-0.0156	-0.0116
lag 7	0.0183	0.0145	0.0533	0.0536	0.0085	0.0066
lag 8	-0.0176	-0.0172	-0.0435	-0.0326	-0.0069	-0.0091
lag 9	0.0323	0.0314	-0.0107	-0.0097	0.0285	0.0207
lag 10	-0.0466	-0.0443	-0.0451	-0.0518	-0.0306	-0.0335
lag 11	-0.0067	-0.0110	0.0740	0.0772	-0.0147	-0.0193
lag 12	-0.0117	-0.0092	0.0275	0.0211	0.0173	0.0148

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, $T = 643$, hence being equal to about 0.00155.

Table 3. Johansen maximum likelihood cointegration procedure

Panel a) S&P 500

LR tests based on the maximum eigenvalue of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r = 1	56.60	14.06
r · 1	r = 2	0.27	3.84

LR tests based on the trace of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r ¸ 1	56.87	15.41
r · 1	r = 2	0.27	3.84

Panel b) NIKKEI 225

LR tests based on the maximum eigenvalue of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r = 1	84.44	14.06
r · 1	r = 2	2.083	3.84

LR tests based on the trace of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r ¸ 1	86.52	15.41
r · 1	r = 2	2.083	3.84

(continued ...)

(... Table 3 continued)

Panel C: FTSE 100

LR tests based on the maximum eigenvalue of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r = 1	60.47	14.06
r · 1	r = 2	0.02	3.84

LR tests based on the trace of the stochastic matrix

H_0	H_1	LR	95% CV
r = 0	r ¸ 1	60.49	15.41
r · 1	r = 2	0.02	3.84

Notes: The model being tested for cointegration is equation (9). H_0 and H_1 denote the null hypothesis and the alternative hypothesis respectively; r denotes the number of cointegrating vectors and the 95% critical values reported in the last column are taken from Osterwald-Lenum (1992).

Table 4. Linearity tests on the residuals from linear VECMs

Panel a) Linear VECM (13) (complete separation)

RESET tests

	S&P 500	NIKKEI 225	FTSE 100
futures equation	0	0	5.90×10^{-2}
spot equation	1.64×10^{-2}	1.29×10^{-3}	7.17×10^{-2}

BDS tests

	v = 2			v = 3		
	& = 0:5	& = 1:0	& = 1:5	& = 0:5	& = 1:0	& = 1:5
			S&P 500			
futures equation	3.02×10^{-6}	8.85×10^{-7}	4.01×10^{-7}	2.10×10^{-7}	1.18×10^{-7}	3.51×10^{-8}
spot equation	1.18×10^{-5}	6.80×10^{-7}	2.66×10^{-6}	2.05×10^{-7}	4.38×10^{-8}	5.79×10^{-8}
			NIKKEI 225			
futures equation	3.13×10^{-8}	2.13×10^{-7}	1.10×10^{-6}	2.77×10^{-14}	4.79×10^{-10}	1.32×10^{-8}
spot equation	2.40×10^{-9}	4.37×10^{-8}	3.86×10^{-7}	2.44×10^{-15}	6.32×10^{-11}	8.42×10^{-9}
			FTSE 100			
futures equation	2.77×10^{-1}	2.77×10^{-1}	2.96×10^{-1}	5.79×10^{-2}	4.86×10^{-2}	3.34×10^{-2}
spot equation	4.30×10^{-1}	3.15×10^{-1}	3.21×10^{-1}	7.14×10^{-2}	4.47×10^{-2}	1.43×10^{-2}

(continued ...)

(... Table 4 continued)

Panel b) Linear VECM (14) (type-B separation)

RESET tests

	S&P 500	NIKKEI 225	FTSE 100
futures equation	7.60×10^{-4}	1.08×10^{-3}	0
spot equation	1.00×10^{-5}	9.47×10^{-3}	0

BDS tests

	$v = 2$			$v = 3$		
	$\varepsilon = 0.5$	$\varepsilon = 1.0$	$\varepsilon = 1.5$	$\varepsilon = 0.5$	$\varepsilon = 1.0$	$\varepsilon = 1.5$
			S&P 500			
futures equation	5.52×10^{-7}	7.42×10^{-7}	1.37×10^{-6}	7.42×10^{-9}	4.31×10^{-8}	9.06×10^{-8}
spot equation	2.09×10^{-6}	2.53×10^{-7}	4.37×10^{-6}	1.59×10^{-8}	5.14×10^{-9}	8.16×10^{-8}
			NIKKEI 225			
futures equation	1.20×10^{-5}	1.99×10^{-5}	3.15×10^{-5}	1.97×10^{-9}	1.71×10^{-7}	1.02×10^{-6}
spot equation	5.36×10^{-7}	6.17×10^{-6}	2.42×10^{-5}	2.93×10^{-10}	5.38×10^{-8}	1.27×10^{-6}
			FTSE 100			
futures equation	2.36×10^{-1}	3.68×10^{-1}	2.77×10^{-1}	3.79×10^{-2}	3.56×10^{-2}	1.14×10^{-2}
spot equation	3.12×10^{-1}	3.08×10^{-1}	4.34×10^{-1}	2.52×10^{-2}	1.55×10^{-2}	9.48×10^{-3}

Notes: Panel a): Under the RESET test statistic, the alternative model involves a higher-order polynomial to represent a different functional form; in the present context we computed the RESET test considering an alternative model with a quadratic and a cubic term under the null of linearity. The RESET test statistic is distributed as $\chi^2(q)$ with q equal to the number of higher-order terms in the alternative model. Panel b): The BDS test statistic tests the null hypothesis that a series is iid against the alternative of a realization from an unspecified nonlinear process (see footnote 15). The critical values, from the normal distribution, are 1.960 and 2.576 at the five percent and one percent nominal levels of significance respectively. For both RESET tests and BDS tests, we report p-values; p-values lower than equal to zero at the 8th decimal point are reported as 0.

Table 5. Likelihood ratio tests

	LR1	LR2
Bivariate VECM for S&P 500	239.49 $\{9.70 \times 10^{-44}\}$	—
Bivariate VECM for NIKKEI 225	222.92 $\{9.29 \times 10^{-42}\}$	—
Bivariate VECM for FTSE 100	155.68 $\{1.47 \times 10^{-26}\}$	—
Multivariate VECM (all indices)	1765.95 $\{7.11 \times 10^{-110}\}$	1658.32 $\{9.38 \times 10^{-95}\}$

Notes: Figures in braces denote $p-$values. LR1 is the likelihood ratio test where the model is not identified under the null due to the existence of nuisance parameters. In this case it tests the null hypothesis of a linear VECM against the alternative hypothesis of an MSIH-VECM with two regimes. LR2 is the likelihood ratio test calculated to test the restrictions in (10). LR2 is distributed as $\chi^2(g)$ where g is the number of restrictions imposed.

Table 6. In-sample performance: coefficients of determination, R^2 and $\overline{R}^2$

	VECM (13)	VECM (14)	MSIH-VECM (15)	MSIH-VECM (16)
		R^2		
S&P 500 f_t	0.038	0.045	0.103	0.660
S&P 500 s_t	0.022	0.029	0.102	0.649
NIKKEI 225 f_t	0.007	0.024	0.018	0.503
NIKKEI 225 s_t	0.006	0.022	0.019	0.454
FTSE 100 f_t	0.011	0.036	0.044	0.550
FTSE 100 s_t	0.004	0.032	0.034	0.553
		$\overline{R}^2$		
S&P 500 f_t	0.037	0.038	0.100	0.260
S&P 500 s_t	0.021	0.025	0.098	0.256
NIKKEI 225 f_t	0.007	0.020	0.017	0.198
NIKKEI 225 s_t	0.006	0.018	0.018	0.179
FTSE 100 f_t	0.011	0.030	0.043	0.217
FTSE 100 s_t	0.004	0.028	0.033	0.218

Notes: R^2 and $\overline{R}^2$ are the coefficient of determination and the adjusted coefficient of determination respectively. The adjusted coefficient of determination is calculated as Krolzig (1997, p. 133-4).

Table 7. Out-of-sample performance: point forecasting

Panel a) R^2 out of sample, mean absolute error and mean square error

	VECM (13)	VECM (14)	MSIH-VECM (15)	MSIH-VECM (16)
		S&P 500		
R^2_{out}	0.02297	0.03263	0.03475	0.15624
MAE	9.83×10^{-3}	9.85×10^{-3}	9.93×10^{-3}	9.62×10^{-3}
MSE	1.48×10^{-4}	1.50×10^{-4}	1.53×10^{-4}	1.47×10^{-4}
		NIKKEI 225		
R^2_{out}	0.00688	0.02867	0.01077	0.19867
MAE	9.01×10^{-3}	8.63×10^{-3}	9.06×10^{-3}	9.21×10^{-3}
MSE	1.17×10^{-4}	1.07×10^{-4}	1.17×10^{-4}	1.36×10^{-4}
		FTSE 100		
R^2_{out}	0.00744	0.03102	0.01165	0.15827
MAE	7.99×10^{-3}	7.70×10^{-3}	7.92×10^{-3}	8.27×10^{-3}
MSE	1.01×10^{-4}	1.01×10^{-4}	9.95×10^{-5}	1.07×10^{-4}

Panel b) Diebold-Mariano test

	$\frac{\text{MSIH-VECM(16)}}{\text{VECM (13)}}$	$\frac{\text{MSIH-VECM(16)}}{\text{VECM(14)}}$	$\frac{\text{MSIH-VECM(16)}}{\text{MSIH-VECM(15)}}$
		S&P 500	
MAE	0.65134	0.61190	0.48856
MSE	0.74570	0.94730	0.81491
		NIKKEI 225	
MAE	0.66788	0.33101	0.75752
MSE	0.15895	0.07743	0.14197
		FTSE 100	
MAE	0.35726	0.09708	0.24916
MSE	0.37056	0.46857	0.29550

Notes: Panel a): R^2_{out} is the out-of-sample coefficient of determination calculated as $R^2_{out} = \sigma^2_{\hat{y}_i}/\sigma^2_y$ where $\sigma^2_{\hat{y}}$ is the variance of the forecast obtained by model i and σ^2_y is the variance of the forecasted variable. MAE and MSE denote the mean absolute error and the mean square error respectively. Panel b): Figures are p-values from executing Diebold-Mariano (1995) test statistics for the null hypothesis that the two competing models i and j have equal point forecast accuracy; $\frac{\text{model i}}{\text{model j}}$ denotes that the two competing models being tested under the Diebold-Mariano statistic are model i and model j. The spectral density of the loss differential function at frequency zero $\hat{f}(0)$ is estimated using the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The p-values were calculated by bootstrap methods using a variant of the procedure suggested by Kilian (1999).

Table 8. Absolute differences in moments

	VECM (13)	VECM (14)	MSIH-VECM (15)	MSIH-VECM (16)
		S&P 500		
μ_1	1.12×10^{-4}	3.73×10^{-4}	5.49×10^{-4}	1.03×10^{-4}
μ_2	1.38×10^{-4}	1.36×10^{-4}	1.36×10^{-4}	1.19×10^{-4}
μ_3	1.28×10^{-7}	1.27×10^{-7}	1.26×10^{-7}	2.46×10^{-7}
μ_4	5.41×10^{-8}	5.40×10^{-4}	5.40×10^{-8}	5.13×10^{-8}
		NIKKEI 225		
μ_1	2.83×10^{-3}	7.81×10^{-4}	2.88×10^{-3}	3.79×10^{-3}
μ_2	1.08×10^{-4}	1.05×10^{-4}	1.07×10^{-4}	8.73×10^{-5}
μ_3	2.53×10^{-7}	2.55×10^{-7}	2.51×10^{-7}	3.05×10^{-7}
μ_4	2.90×10^{-8}	2.90×10^{-8}	2.90×10^{-8}	2.72×10^{-8}
		FTSE 100		
μ_1	9.98×10^{-4}	1.05×10^{-3}	1.04×10^{-3}	4.50×10^{-5}
μ_2	9.99×10^{-5}	9.75×10^{-5}	9.95×10^{-5}	9.48×10^{-5}
μ_3	6.31×10^{-7}	6.33×10^{-7}	6.29×10^{-7}	6.25×10^{-7}
μ_4	3.30×10^{-8}	3.30×10^{-8}	3.30×10^{-8}	3.27×10^{-8}

Notes: μ_1, μ_2, μ_3, μ_3 are the absolute differences between the means, variances, third and fourth central moments of the estimated (linear and nonlinear) VECM models and the corresponding moments of actual data.

Table 9. Out-of-sample performance: density forecasting

$\frac{MSIH-VECM(16)}{VECM(13)}$	$\frac{MSIH-VECM(16)}{VECM(14)}$	$\frac{MSIH-VECM(16)}{MSIH-VECM(15)}$
	S&P 500	
$\begin{bmatrix}4.8194\\1.43\times10^{-6}\end{bmatrix}$	$\begin{bmatrix}4.3966\\1.09\times10^{-5}\end{bmatrix}$	$\begin{bmatrix}3.9948\\6.47\times10^{-5}\end{bmatrix}$
	NIKKEI 225	
$\begin{bmatrix}5.6034\\2.10\times10^{-8}\end{bmatrix}$	$\begin{bmatrix}2.0967\\3.60\times10^{-2}\end{bmatrix}$	$\begin{bmatrix}5.4505\\5.02\times10^{-8}\end{bmatrix}$
	FTSE 100	
$\begin{bmatrix}4.7508\\2.02\times10^{-6}\end{bmatrix}$	$\begin{bmatrix}1.8190\\6.89\times10^{-2}\end{bmatrix}$	$\begin{bmatrix}4.2595\\2.04\times10^{-5}\end{bmatrix}$

Notes: Figures denote the ´-test statistic for equal density forecast accuracy and corresponding $p-$values (in brackets), constructed as described in the text. The null hypothesis is that the two competing models i and j have equal density forecast accuracy; $\frac{\text{model } i}{\text{model } j}$ denotes that the two competing models being tested under the ´-test statistic are model i and model j. The integrated square differences were calculated using the Gaussian kernel estimator and setting the smoothing parameter according to the Silverman's (1986) rule. The number of bootstrap replications $B = 100$. The test is distributed as $N(0;1)$ under the null (see Appendix B).

A The transition matrix of the MSIH-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 $z_t \in \{1, \ldots, M\}$ governed by the transition probabilities $p_{ij} = \Pr(z_t = j \mid z_{t-1} = i)$, 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, MSIH-VECM (16)), we need to specify the joint process governing the transitional dynamics of the whole system. Define z_t^{SP}, z_t^{NK} and z_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 z_t^{χ} would be an outcome of a Markov chain with transition probabilities p_{ij}^{χ} where z_t^{χ} is independent of z_t^{ψ} with $\psi \neq \chi$ for any t. In order to analyze the whole dynamics of the MSIH-VECM (16) we construct the following latent variable

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

Under this formalization the latent variable ξ_t governing the transitional dynamics of the whole system MSIH-VECM (16) follows an 8-state Markov chain whose transition probabilities can be easily calculated from the probabilities of the chain governing z_t^{SP}, z_t^{NK} and z_t^{FT}. For example:

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

B The η-test for equal density forecast accuracy

This appendix briefly outlines the derivation of the η-test statistic for the null hypothesis of equal density forecast accuracy. Consider two series of forecasts, say $\{\hat{y}_{1t}\}_{t=1}^{T_1}$ and $\{\hat{y}_{2t}\}_{t=1}^{T_2}$, obtained from two competing models, say M_1 and M_2. Let $f(y)$ be the probability density function of the variable y_t over the period $t = 1, \ldots, T$, and $g_1(y)$ and $g_2(y)$ be the probability density functions of the two forecast series $\{\hat{y}_{1t}\}_{t=1}^{T_1}$ and $\{\hat{y}_{2t}\}_{t=1}^{T_2}$ respectively.[30] Assume that $f(y)$, $g_1(y)$ and $g_2(y)$ are associated with distribution functions F, G_1 and G_2 respectively, and F, G_1 and G_2 are absolutely continuous with respect to the Lesbegue measure in R^p.

We are interested in testing the null hypothesis of equidistance of the probability densities $g_1(y)$ and $g_2(y)$ from $f(y)$, that is

$$H_0 : \text{dist}[f(y); g_1(y)] = \text{dist}[f(y); g_2(y)]; \tag{B1}$$

where the operator dist denotes a generic measure of distance:

A common measure of global closeness between two functions is the integrated square difference (ISD) (e.g. see Pagan and Ullah, 1999):

$$ISD = \int [\phi(x) - \gamma(x)]^2 dx; \tag{B2}$$

where $\phi(\cdot)$ and $\gamma(\cdot)$ denote probability density functions; $ISD \geq 0$, and $ISD = 0$ only if $\phi(x) = \gamma(x)$. Using (B2) we can rewrite the null hypothesis H_0 in (B1) as follows:

$$\begin{aligned} H_0 \quad &: \int [f(y) - g_1(y)]^2 dy = \int [f(y) - g_2(y)]^2 dy \\ &: ISD_1 - ISD_2 = 0. \end{aligned} \tag{B3}$$

Under (B3) the null hypothesis of equal density forecast accuracy of models M_1 and M_2 is written as the null hypothesis of equality of two integrated square differences or, equivalently, as the null hypothesis that the difference between two integrated square differences is zero.

With observations $\{y_t\}_{t=1}^{T}$, $\{\hat{y}_{1t}\}_{t=1}^{T}$ and $\{\hat{y}_{2t}\}_{t=1}^{T}$ we can consistently estimate the unknown functions $f(y)$, $g_1(y)$ and $g_2(y)$ using kernel estimation to obtain:

$$\hat{f}(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_i - y}{h}\right) \tag{B4}$$

$$\hat{g}_1(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_{1i} - y}{h}\right) \tag{B5}$$

$$\hat{g}_2(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_{2i} - y}{h}\right) \tag{B6}$$

where $K(\cdot)$ is the kernel function and h is the smoothing parameter. Using (B4)-(B6) we can then obtain a consistent estimate of the integrated square differences ISD_1 and ISD_2. Define $d = \widehat{ISD}_1 - \widehat{ISD}_2$ as the estimated relative distance between the probability density functions. In order to test for the statistical significance of d, the next step is to calculate a confidence interval for d.

[30] For simplicity and for clarity of exposition, throughout this section, we consider the case where $T_1 = T_2 = T$, although the results derived below can be easily extended to the more general case where $T_1 \neq T_2 \neq T$.

In the spirit of the analysis of Hall (1992), define $\left\{y_i^j\right\}_{i=1}^T$, $\left\{\hat{y}_{1i}^j\right\}_{i=1}^T$, $\left\{\hat{y}_{2i}^j\right\}_{i=1}^T$ as the $j-$th resample of the original data $\{y_t\}_{t=1}^T$, $\{\hat{y}_{1t}\}_{t=1}^T$, $\{\hat{y}_{2t}\}_{t=1}^T$, drawn randomly with replacement. From these resamples it is possible to obtain consistent bootstrap estimates of the density functions $\hat{f}^j(y)$, $\hat{g}_1^j(y)$, $\hat{g}_2^j(y)$ and, consequently, of $d^j = \widehat{ISD}_1^j - \widehat{ISD}_2^j$.

Consider a sample path $\left\{d^j\right\}_{j=1}^B$, where B is the number of bootstrap replications. Under general conditions[31], we have:

$$\sqrt{B}\left(\bar{d} - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right); \tag{B7}$$

where

$$\bar{d} = \frac{1}{B}\sum_{j=1}^{B} d^j = \frac{1}{B}\sum_{j=1}^{B}\left(\widehat{ISD}_1^j - \widehat{ISD}_2^j\right) \tag{B8}$$

is the average difference of the estimated distances over B bootstrap replications. Because in large samples the average difference $\bar{d}$ is approximately normally distributed with mean μ and variance σ^2/B, the large sample statistic for testing the null hypothesis that models M_1 and M_2 have equal density forecast accuracy is:

$$\eta = \frac{\bar{d}}{\sqrt{\frac{\hat{\sigma}^2}{B}}} \xrightarrow{d} N(0, 1); \tag{B9}$$

where $\hat{\sigma}^2$ is a consistent estimate of σ^2. [32]

As mentioned in the main text, this test statistic may be seen as the analogue of the test of Diebold and Mariano (1995) in the context of density forecasting. Sarno and Valente (2001), using Monte Carlo methods designed to investigate the size and power properties of this test statistic, show that the η-test has satisfactory empirical size and power properties in finite sample in a variety of circumstances with a number of boostrap replications equal to 100 or so.

[31] See Kendall and Stuart (1976, Ch. 11).

[32] On the consistency of the bootstrap estimates of σ^2 in this context see Hall (1992) and Mammen (1992).

References

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 8207.

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.

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

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.

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.

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.

Hall, P. (1992), \E®ect of Bias Estimation on Coverage Accuracy of Bootstrap Con¯dence Intervals for a Probability Density" Annals of Statistics, 20, 675-694.

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. (1992), \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.

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.

Johansen, S. (1995), Likelihood-based Inference in Cointegrated VAR Models. Oxford: Oxford University Press.

Karatzas, I. and Shreve, S.E. (1998), Methods of Mathematical Finance, NY: Springer.

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 Gri±n 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. (2001), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?", Centre for Economic Policy Research Discussion Paper No. 3024.

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.

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 Autoregressive Processes," Nuffield College and Department of Economics, University of Oxford, mimeo.

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

Lee, B.-S. (1998), "Permanent, Temporary, and Non-fundamental Components of Stock Prices," Journal of Financial and Quantitative Analysis, 33, 1-32.

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.

Mammen, E. (1992), When Does Bootstrap Work: Asymptotic Results and Simulations, Lecture Notes in Statistics, 77, Berlin: Springer-Verlag.

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.

Pagan, A. and Ullah, A. (1999), Nonparametric Econometrics, Cambridge, New York and Melbourne: Cambridge University Press.

Pesaran, M.H. and Skouras, S. (2001), "Decision-based Methods for Forecast Evaluation," in Clements, M.P. and Hendry, D.F. (eds.), A Companion to Economic Forecasting, Oxford: Blackwell, forthcoming.

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.

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.

Ryden, T., TerÄsvirta, T. and Asbrink, 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 In¯nite Order Vector Autoregressive Processes," Journal of Econometrics, 81, 93-126.

Sarno, L. and Valente, G. (2001), \Comparing the Accuracy of Density Forecasts from Competing Models," University of Warwick, mimeo.

Silverman, B.W. (1986), Density Estimation for Statistics and Data Analysis, New York: Chapman and Hall.

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.

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 Di®erences of Equivalent Assets," Mathematical Finance, 4, 205-221.