

A new econometric model of index arbitrage*

Nicholas Taylor[†]
Cardiff University

January 12, 2004

Abstract

This paper introduces a new econometric model of the mispricing associated with (contemporaneous) differences between spot and futures prices. Like existing models, this model assumes that the level of arbitrage activity is positively related to the magnitude of absolute mispricing. However, unlike existing models, the new model assumes that a parameter governing a key feature of this relationship is exogenously determined and, as such, varies over time. Specifically, a smooth transition model of mispricing, and a corresponding Lagrange multiplier (LM) linearity test, is introduced that allows the degree of ‘smoothness’ in the transition function to be determined by a set of explanatory variables. Using high frequency data from the S&P 500 spot and futures market, the results show that the nature of the non-linearity in mispricing corresponds to smoothness that varies (in a periodic fashion) over the trading day. This is evinced by the LM test results and by the superior fit of the new model of mispricing, in comparison to the results based on existing econometric models of mispricing. Finally, the observed periodicity in arbitrageur behavior indicates that arbitrageurs prefer to trade at the beginning, rather than at the end, of the trading day – a result that contradicts the findings obtained when using existing econometric models of mispricing.

Key Words: Index arbitrage, smooth transition, intraday periodicity.

JEL Classification Code: C22; C41; G14.

*I thank Cardiff Business School for funding received while undertaking this paper, and John Huang of *Tick Data Inc.* for generously supplying the data used in this paper.

[†]Department of Accounting and Finance, Cardiff University, Cardiff CF10 3EU, United Kingdom. Email: taylorn@cardiff.ac.uk.

1 Introduction

Issues concerning the relationship between stock index levels and stock index futures prices have generated considerable debate within the finance literature. While an array of issues have been debated over the years, it is those pertaining to the observed dynamics of mispricing that have attracted most attention.¹ The most recent wave of interest has been on the importance of certain features of the arbitrage process in determining mispricing dynamics. Specifically, behavioral aspects of the arbitrage process have been shown to give rise to non-linear time-dependence in mispricing. It is to this tranche of the literature that this paper contributes. In particular, a new econometric model of mispricing is introduced that is compatible with arbitrage being costly and risky (referred to as limited arbitrage).² However, unlike existing models, this model allows these behavioral aspects of arbitrage to be exogenously determined and, hence, to vary over time – a feature commonly observed in financial markets. In using this model, new evidence of the nature of arbitrageur behavior is provided.

Various econometric models of mispricing have been employed in the literature. The exact specification of mispricing used has largely been a function of the assumed extent, and nature, of the arbitrageur behavior underlying the mispricing. Under the classic assumption of market efficiency, early studies claim that unlimited (costless and riskless) arbitrage activity generates a linear structure in mispricing dynamics; see MacKinlay and Ramaswamy (1988), Yadav and Pope (1990), and Lim (1990), for empirical evidence from a variety of markets. However, Harris (1989), Kleidon (1992), and Miller, Muthuswamy and Whaley (1994) question this evidence, and argue instead that underlying market microstructure frictions, particularly non-trading in the stocks within the underlying index, determine the linear evolution of mispricing over time. More recently, Garrett and Taylor (2001) and Tse (2001) successfully incorporate behavioral aspects of the arbitrage process into non-linear models of mispricing, and demonstrate that market microstructure issues merely obfuscate the analysis. In particular, they show that, while

¹Mispricing is defined as the difference between the theoretical, or fair, price of the futures contract, and the market price of the same futures contract. By contrast, the basis is defined as the difference between the market price of the futures contract and the price (level) of the underlying asset (index).

²See Barberis and Thaler (2003) for a survey of behavioral finance that includes a detailed discussion of limited arbitrage.

market microstructure plays a minor role, it is limited arbitrage activity that determines the vast majority of the observed mispricing dynamics. Given the success of models that more closely match the behavior of arbitrageurs, this paper extends the realism of the behavioral assumptions underlying the arbitrage process to produce an econometric model that more accurately represents mispricing dynamics.

There are two existing models of mispricing that place a strong emphasis on behavioral aspects of the arbitrage process: those that assume time-invariant homogeneous arbitrageur behavior, and those that assume time-invariant heterogeneous arbitrageur behavior. Both classes of model are based on the premise that arbitrage activity will only occur when the profit opportunity is sufficiently large. However, they differ in that the first class of model assumes that all arbitrageurs agree on what constitutes a ‘sufficiently large’ profit opportunity; whereas the second class assumes that arbitrageurs may differ in their assessment of the profit opportunity. Crucially, both classes of model assume that this assessment process is not a function of exogenous factors and, hence, does not vary over time. This would appear to be a rather restrictive assumption given that the quality of the profit opportunity is highly dependent upon, *inter alia*, implementation costs and/or the risk of the opportunity itself – both of which one would expect to vary over time. Therefore, motivated by this reasoning, time-varying heterogeneous arbitrageur behavior is assumed, and a third class of model is introduced that allows the evaluation of the quality of the profit opportunity to be exogenously determined and, hence, to vary over time.

Under the assumption of time-invariant homogeneous arbitrageur behavior, a threshold model can be used to capture the non-linear dynamics of mispricing (see, e.g., Yadav, Pope and Paudyal, 1994, Dwyer, Locke and Yu, 1996, Martens, Kofman and Vorst, 1998, Forbes, Kalb and Kofman, 1999, and Garrett and Taylor, 2001). In the simplest version of the threshold model, mispricing can fall within two regimes: it can be within the non-profitable bound associated with an arbitrage transaction, in which case no arbitrage trading occurs, or it can be outside it, in which case arbitrage activity takes place. As predicted by theory, the general conclusion of these studies is that no arbitrage trading in the spot or futures markets occurs

(mispricing follows a non-stationary process) when previous mispricing lies inside the bound, but that offsetting arbitrage positions in the spot and futures are taken (mispricing follows a stationary process) when previous mispricing lies outside the bound.

Similar conclusions are obtained when one makes the assumption of time-invariant heterogeneous arbitrageur behavior. In this case a smooth transition model can be applied to the mispricing series (see, e.g., Taylor et al., 2000, and Tse, 2001). This model makes use of a smooth transition function to measure the extent to which mispricing represents a profitable opportunity, where the ‘smoothness’ of the function is assumed to be the result of the aggregation of arbitrageurs who possess different opinions on whether mispricing can be traded with profit. Though this model represents an improvement over the simple threshold model, it fails to capture salient features of the index arbitrage market; specifically, the propensity-to-arbitrage (for a given level of mispricing) in this model, as measured by the shape of the transition function (smoothness), is assumed to remain constant over time. However, it is well documented that there are time-varying factors, such as implementation costs and/or risks pertaining to the arbitrage process, that are likely to affect this propensity. Therefore, it seems reasonable to expect the propensity-to-arbitrage to vary over time. It is for this reason that a new econometric model is introduced that relaxes the strong assumption that the propensity-to-arbitrage is time-invariant.

Using intraday data on the S&P 500 stock index and associated index futures contracts, we examine the quality of the model of mispricing introduced in this paper, in comparison to existing models of mispricing; namely, the threshold and smooth transition models. This quality is assessed by conventional model fit metrics, and by carrying out a new Lagrange multiplier (LM) test for linearity in mispricing (against the alternative of non-linearity as implied by the new model). This paper also innovates by using data that more accurately matches the behavior of arbitrageurs (and mitigates certain market microstructure effects). In particular, the prices of the S&P 500 iShares exchange traded fund (ETF) are used to measure the levels of the index. Also, we examine the robustness of the results by using two different types of S&P 500 futures contracts: regular S&P 500 futures contracts and S&P 500 E-mini futures contracts. The use

of the latter series is motivated by the results obtained by Hasbrouck (2003), who shows that most S&P 500 price discovery occurs in the E-mini futures market. To anticipate the results, we find that an LM test for linearity, against an alternative that corresponds to a model that allows the shape of the transition function (smoothness) to vary over the trading day, is categorically rejected. Moreover, this model is significantly more accurate at representing the dynamics of mispricing, in comparison to existing mispricing models. Finally, the intraday periodicity in the smoothness parameter throws light on behavioral aspects of the arbitrage process in the S&P 500 index arbitrage market. Interesting, the observed periodicity shows that arbitrageurs prefer to trade at the beginning, rather than at the end, of the trading day – a result that contradicts the findings obtained when using existing econometric models of mispricing.

The rest of the paper is organized as follows. In the next section we outline the cost-of-carry model of stock index futures pricing. Section 3 contains detailed descriptions of the threshold model, the smooth transition model, and the new model of mispricing introduced in this paper. In Section 4, a test for linearity (against non-linearity based on the new model of mispricing) is derived and examined. The penultimate section contains a discussion of the data used, and presents results pertaining to various tests and econometric models of mispricing dynamics. Finally, Section 6 offers some concluding remarks.

2 An economic model of index arbitrage

The contemporaneous relationship between spot and forward prices can be described by the cost-of-carry model. This model is also capable of describing the relationship between spot and futures prices providing that the term structure of interest rates is flat and constant. Under this assumption, and in the absence of arbitrage opportunities, we have

$$\tilde{F}_{t, T_m} = S_t e^{r(T_m - t)} - \sum_{h=1}^H D_h e^{r(T_m - \tau_h)}, \quad (1)$$

where \tilde{F}_{t,T_m} is the theoretical (or fair) stock index futures price observed at time t for delivery at time T_m , S_t is the level of the index, r is the risk-free continuous interest rate applicable over the contract life, $(T_m - t)$ is the time-to-maturity of the futures contract, and D_h is the expected cash dividend paid at time τ_h , where $t < \tau_h \leq T_m$.

The economic model given by (1) can be used to construct an arbitrage decision rule. Specifically, if the contract is held to maturity then, in the presence of proportional (and symmetric) implementation costs, $\kappa > 0$, arbitrage activity will take place when one of the following conditions holds:

$$\frac{F_{t,T_m}}{\tilde{F}_{t,T_m}} < 1 - \kappa, \quad (2a)$$

$$\frac{F_{t,T_m}}{\tilde{F}_{t,T_m}} > 1 + \kappa, \quad (2b)$$

where κ equals the sum of commissions, bid-ask spreads, the price impact costs associated with arbitrage trades, *and* the opportunity costs associated with constrained arbitrageurs not being able to trade.³ The constitution of κ is further complicated by the fact that arbitrageurs can, and do, unwind their spot and futures positions before maturity (see, inter alia, Sofianos, 1993, and Neal, 1996, for empirical evidence). Brennan and Schwartz (1990) argue that this ability of arbitrageurs to close out their positions prior to maturity, leads to a narrowing of the arbitrage bound. Indeed, they argue that arbitrage will only take place when absolute mispricing exceeds implementation costs *minus* the value of the option to liquidate the position early.

Apart from narrowing the arbitrage bound, early liquidation also introduces the possibility that κ may also measure risk factors pertaining to the arbitrage process. Specifically, there is a risk (referred to as noise trader risk) that the difference between F_{t,T_m} and \tilde{F}_{t,T_m} may increase in the short-run.⁴ Such risk is not important if the position can be held until maturity. However, arbitrage is a highly specialized activity requiring vast amounts of external capital

³Equations (2a) and (2b) assume that the arbitrage bound is symmetric. However, it is often argued that short-selling restrictions on the stocks within the index will result in an asymmetric arbitrage bound; see Fung and Draper (1999) for a discussion of these issues in the context of index arbitrage, and D'Avolio (2002) for a more general discussion of short-selling restrictions. Whether the symmetric arbitrage bound assumption is valid is examined later in the paper.

⁴See De Long et al. (1990) for a discussion of the impact of noise trader risk on financial markets.

for successful execution. As such, Shleifer and Vishny (1997) argue that, as an arbitrageur is essentially an agent managing the money of a less well informed principal, her positions may be closed prematurely (by the principal) to avoid large short-run losses. As such, initial absolute differences between F_{t,T_m} and \tilde{F}_{t,T_m} may need to be higher to avoid noise trader risk – implying that κ will be higher in the presence of such risk.

There is also a risk associated with the time taken for arbitrageurs to adopt appropriate positions in the stocks within the index and the associated futures contracts. To incorporate this feature of the arbitrage process, the arbitrage opportunity available at $t - d$ is used as the expectation of mispricing at t (conditional on information available at $t - d$, which is denoted Ω_{t-d}). Therefore, providing κ is small, (2a) and (2b) can be expressed as

$$|y_{t-d}| > \kappa, \tag{3}$$

where $y_t = \ln F_{t,T_m} - \ln \tilde{F}_{t,T_m}$ is the definition of mispricing used in this paper, $y_{t-d} = \mathbf{E}[y_t | \Omega_{t-d}]$, $\mathbf{E}[\cdot]$ is an expectation operator, and d is the delay inherent in the arbitrage process. Under the assumption of time-invariant homogeneous arbitrageur behavior, all arbitrageurs face the same arbitrage condition as given by (3) – an assumption that provides the foundation for the threshold econometric model used to describe arbitrage behavior.

The validity of the threshold model relies heavily on the assumption of homogeneous arbitrageur behavior. However, this assumption is unrealistic with respect to the above arbitrage condition as it is highly likely that each arbitrageur will face different implementation costs (net of the early liquidation option value) and noise trader risk. For instance, member firm arbitrageurs pay no commission on arbitrage trades, while all other arbitrageurs pay pre-arranged (non-zero) fees. Similarly, arbitrageurs will face different levels of noise trader risk. An obvious case will be the difference in noise trader risk faced by successful (low risk) and unsuccessful (high risk) arbitrageurs. To accommodate such time-invariant heterogeneous arbitrageur behavior, we

assume that each arbitrageur faces an arbitrage condition of the form,

$$|y_{t-d}| > \kappa_i, \tag{4}$$

where κ_i is the sum of the implementation costs (net of the early liquidation option value) and noise trader risk faced by the i th arbitrageur. As will be described in the following section, aggregation of this condition over all arbitrageurs motivates use of the smooth transition econometric model.

While the arbitrage condition has been adapted for heterogeneity in arbitrage behavior, no attempt has been made to introduce heterogeneity over time. This is a serious omission given that it is highly likely that the size of the arbitrage bound (inclusive of risks, etc.) is likely to vary over time. More specifically, factors affecting this bound, such as bid-ask spreads and price risk, exhibit a strong intraday periodicity; see, *inter alia*, Wood, McInish and Ord (1985), Baillie and Bollerslev (1991), Harvey and Haung (1991), McInish and Wood (1992), Werner and Kleidon (1996), and Tse (1999). Given the likely time variation in the size of the arbitrage bound, we assume that the i th arbitrageur forms an expectation of the size of this bound (conditional upon information available at a point in time that is consistent with any delays inherent in the arbitrage process), and trades accordingly. Such reasoning implies that the arbitrage condition in (3), augmented by the assumption of time-varying heterogeneous arbitrageur behaviour, is given by

$$|y_{t-d}| > \mathbf{E}[\kappa_{i,t} | \Omega_{t-d}], \tag{5}$$

where $\kappa_{i,t}$ is the sum of the implementation costs (net of the early liquidation option value) and noise trader risk faced by the i th arbitrageur and time t . As will be described in the following section, the econometric model introduced in this paper becomes an appropriate mispricing model under an aggregated version of this arbitrage condition.

3 Econometric models of index arbitrage

In this section, we outline the general motivation lying behind econometric models of mispricing, and provide specific details of the econometric models of mispricing associated with the arbitrage conditions given by (3), (4), and (5); namely, the threshold, the smooth transition, and an augmented smooth transition autoregressive models. In addition, an LM test for linearity, against the alternative of non-linearity as implied by the augmented smooth transition autoregressive model, is derived.

3.1 Motivation

The econometric models used in this paper are motivated by examining the possible shortcomings of competing models of arbitrageur behavior. The main competition comes from studies that adopt a non-econometric approach to this behavior (see, e.g., Bailey, 1989, Yadav and Pope, 1990, and Sofianos, 1993). In these studies, arbitrage activity is examined by focusing on the relationship between mispricing and an *assumed* arbitrage bound, without using an econometric model of mispricing. As with studies based on econometric models, non-econometric studies assume that arbitrageur activity occurs when absolute mispricing exceeds the arbitrage bound. However, the latter studies differ in that they rely on a single (or small number of different) *constructed* arbitrage bound(s) – a difficult task given the various components of this bound.

The above issue provides the motivation for use of econometric models when analyzing arbitrageur behavior. Specifically, by using econometric models one can avoid the need to construct the arbitrage bound. This is achieved by assuming that the parameters of the model measure the magnitude of the arbitrage bound. Moreover, the analysis can be further extended by letting the magnitude of the arbitrage bound vary over arbitrageurs and time – features that are almost impossible to capture using non-econometric approaches.

3.2 The threshold autoregressive model

The key feature key of this econometric model is the differential treatment of mispricing inside and outside of the arbitrage bound. If mispricing is expected to remain within the arbitrage bound then it will follow a unit root process. This is because there will be no arbitrage activity driving prices back into equilibrium. However, if expected mispricing goes above or below the arbitrage bound then mispricing will follow a stationary autoregressive (AR) process, as arbitrage activity forces prices back into equilibrium. For instance, if expected mispricing goes above the upper limit of the bound then arbitrageurs will sell the futures contract and buy stocks within the index, hence forcing futures prices down and spot prices up until mispricing no longer violates the arbitrage bound.

The simplest way to model this dynamic price behavior is by means of the threshold autoregressive (TAR) model (see Tong, 1983, 1990, Tong and Lim, 1980, Tsay, 1989, and Hansen, 1996, 1997, 1999, 2000, for theoretical treatments of this model, and Yadav, Pope and Paudyal, 1994, Forbes, Kalb and Kofman, 1999, and Garrett and Taylor, 2001, for applications to stock index futures mispricing).⁵ Formally, y_{t-d} is continuously valued so that partitioning the real line defines the number of distinct regimes into which y_{t-d} can fall. In the current application, the arbitrage condition given by (3) implies that y_{t-d} (the measure of expected mispricing used in this paper) can enter one of two regimes: the non-profit opportunity regime or the profit opportunity regime. The process is in the former regime when $-\kappa \leq y_{t-d} \leq \kappa$, is in the latter regime when this condition does not hold, and is assumed to evolve as follows:

$$y_t = \varphi'_1 \mathbf{x}_t (1 - G(y_{t-d}; \kappa)) + \varphi'_2 \mathbf{x}_t G(y_{t-d}; \kappa) + \epsilon_t, \quad (6)$$

where $\varphi_j = [\varphi_{0,j}, \varphi_{1,j}, \dots, \varphi_{p,j}]'$, $j \in \{1, 2\}$, $\mathbf{x}_t = [1, \tilde{\mathbf{x}}_t']'$, $\tilde{\mathbf{x}}_t = [y_{t-1}, \dots, y_{t-p}]'$, $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$, $t \in \{1, \dots, T\}$, and $G(y_{t-d}; \kappa)$ is a discrete transition function (often referred to as the threshold

⁵The analysis in this paper is based on mispricing following an AR process within a number of the pre-selected regimes. It is, however, possible to model mispricing in the context of an error-correction framework (see, e.g., Dwyer, Locke and Yu, 1996, Martens, Kofman and Vorst, 1998, Taylor et al., 2000, and Tse, 2001). Though not explicitly derived, it is possible to augment the model introduced in this paper within the error-correction framework.

function) that equals unity if the arbitrage condition given by (3) holds, and zero otherwise. This model (henceforth denoted M1) provides values of $G(y_{t-d}; \kappa)$ that can be interpreted as a measure of the level of arbitrage activity.

3.3 The smooth transition autoregressive model

Use of the smooth transition autoregressive (STAR) model in this paper relies on an explicit assumption concerning the aggregation of the arbitrage condition, given by (4), over all arbitrageurs. Following Taylor et al. (2000) and Tse (2001), it is assumed that this aggregation process gives rise to an exponential smooth transition function (see Granger and Teräsvirta, 1993, and Teräsvirta, 1994, for detailed descriptions of this particular version of the STAR model).⁶ Use of this transition function implies that mispricing evolves as follows:

$$y_t = \varphi_1' \mathbf{x}_t (1 - G(y_{t-d}; \gamma, c)) + \varphi_2' \mathbf{x}_t G(y_{t-d}; \gamma, c) + \epsilon_t, \quad (7)$$

where $G(y_{t-d}; \gamma, c)$ is a continuous transition function that is bounded between zero and unity, and is assumed to be given by

$$G(y_{t-d}; \gamma, c) = 1 - \exp(-\gamma(y_{t-d} - c)^2), \quad \gamma > 0, \quad (8)$$

where γ and c are the slope (or smoothness) and location parameters associated with the transition variable, y_{t-d} , respectively.⁷

As with the threshold function used in M1, the above smooth transition function is bounded between zero and unity, and is positively related to absolute mispricing. Furthermore, these extreme values take on a similar interpretation to that taken in M1: a function value of zero implies that no arbitrageurs will trade, while a value of unity implies that all arbitrageurs will trade. As such, this model (henceforth denoted M2) also provides values of the transition

⁶In should noted that, although this paper uses the exponential transition function, it is a trivial matter to alter the subsequent analysis to take account of alternative transition functions such as the logistic function.

⁷The inclusion of exogenous variables in (7) is permitted without complicating the estimation process; see Teräsvirta (1998) for more details of STAR models with exogenous variables included. However, for simplicity, such variables are not included in the present study.

function that can be interpreted as a measure of the level of arbitrage activity. However, unlike the threshold function, this level changes from zero to unity in a smooth fashion, with the rate of transition from zero to unity measured by the smoothness parameter, γ . As higher (lower) values of γ indicate a greater (lesser) desire to arbitrage, *ceteris paribus*, this parameter provides a measure of the propensity-to-arbitrage.

3.4 An augmented smooth transition autoregressive model

The above specification assumes that the smoothness parameter, γ , is time-invariant. However, in the presence of the arbitrage condition given by (5), this assumption is unlikely to hold. The reasoning behind the relaxation of this restriction proceeds as follows: First, the expectation operator on the RHS of (5) implies a functional dependence between $\kappa_{i,t}$ and Ω_{t-d} . Second, aggregation of this condition over all arbitrageurs, in turn, implies that the smoothness parameter, γ_t , will also be functionally dependent on Ω_{t-d} and will, thus, vary over time. Consequently, using reasoning analogous to that used for the STAR model, mispricing is assumed to evolve as follows:

$$y_t = \varphi_1' \mathbf{x}_t (1 - G(y_{t-d}; \gamma_t, c)) + \varphi_2' \mathbf{x}_t G(y_{t-d}; \gamma_t, c) + \epsilon_t, \quad (9)$$

where $G(y_{t-d}; \gamma_t, c)$ is a continuous transition function that is bounded between zero and unity, and is assumed to be given by

$$G(y_{t-d}; \gamma_t, c) = 1 - \exp(-\gamma_t (y_{t-d} - c)^2), \quad \gamma_t > 0, \quad (10)$$

where $\gamma_t = f(\Omega_{t-d})$, with $f(\cdot)$ denoting an unspecified functional dependence. As with previous models, this transition function, and its associated smoothness parameter, can be interpreted as measures of the level of arbitrage activity and the propensity-to-arbitrage, respectively.

To make this model operational we need to specify the nature of the functional dependence between γ_t and Ω_{t-d} , and require knowledge on the contents of Ω_{t-d} . As there is no theoretically

motivated specification for the nature of the functional dependence, we assume (for simplicity) that γ_t is linearly dependent on Ω_{t-d} , and that Ω_{t-d} consists of r independent variables, denoted \mathbf{Z}_{t-d} , hence

$$\gamma_t = \theta_0 + \boldsymbol{\theta}'\mathbf{Z}_{t-d}, \quad (11)$$

where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_r]'$, and $\mathbf{Z}_{t-d} = [Z_{1,t-d}, \dots, Z_{r,t-d}]'$.⁸

The selection of the variables in \mathbf{Z}_{t-d} is, at first blush, a difficult task given the vast number of factors affecting arbitrage bounds. However, this problem can be considerably simplified by noting that these factors are likely to process a strong intraday periodic component. For instance, it is well established (both empirically and theoretically) that bid-ask spreads in spot and futures markets follow an intraday U-shaped pattern; see McNish and Wood (1992) and Tse (1999) for empirical evidence from these two markets, respectively. As such, the approach taken in this paper is to allow γ_t to be determined by a set of intraday periodic components.

One possible model allows γ_t to take a different value during each t_s (intraday) time period within the periodic cycle (the trading day) of length T_s , where $t_s \in \{1, 2, \dots, T_s - 1, T_s\}$, and t_s and t are related by the function $f_s(t)$.⁹ Under this assumption,

$$\boldsymbol{\theta}' = [\theta_2, \dots, \theta_{t_s}, \dots, \theta_{T_s}], \quad (12a)$$

$$\mathbf{Z}'_t = [D_{2,t}, \dots, D_{t_s,t}, \dots, D_{T_s,t}], \quad (12b)$$

and the dummy variable, $D_{t_s,t}$, takes a value of unity if the current observation is in the t_s th stage of the periodic cycle, and a value of zero otherwise.¹⁰ A problem with this model is that

⁸There is a related class of econometric model already existing in the literature. In particular, Lundbergh, Teräsvirta and van Dijk (2003) introduce a time-varying STAR (TV-STAR) model that allows an addition time-dependent transition function in the STAR model that, in turn, permits a smooth change in the parameters of the STAR model. While this model is successfully applied to several macroeconomic (and other miscellaneous) time series, application to mispricing is inappropriate on two counts: First, economic theory predicts that time-variation will only affect the smoothness parameter. This prediction is driven by the action of individual arbitrageurs who continuously change their expectation of the future magnitude of the arbitrage bound. Second, time-variation is likely to be far from smooth, with arbitrageurs changing these expectations at very high frequencies. For these reasons, this model is not considered in this paper.

⁹The function $f_s(t)$ is defined such that

$$(t_s, t) = (1, 1), (2, 2), \dots, (T_s - 1, T_s - 1), (T_s, T_s), (1, T_s + 1), (2, T_s + 2), \dots, (T_s, T).$$

¹⁰Such dummy variables have already been proposed in the context of a periodic STAR model. In particular,

a large number of coefficients are required if there are many time periods within each periodic cycle (i.e., T_s is large). It is possible to overcome this problem by selecting dummy variables that span more than one time period. However, this assumes that γ_t is constant within the time period covered by the dummy variable and then changes abruptly whenever a new time period is entered.

To overcome the above problems, γ_t is assumed to be a function of the flexible Fourier form (FFF), and is given by (11) with

$$\boldsymbol{\theta}' = [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_m, \dots, \boldsymbol{\theta}'_M], \quad (13a)$$

$$\boldsymbol{\theta}'_m = [\theta_{1,m}, \theta_{2,m}], \quad (13b)$$

$$\mathbf{Z}'_t = [\mathbf{Z}'_{1,t}, \dots, \mathbf{Z}'_{m,t}, \dots, \mathbf{Z}'_{M,t}], \quad (13c)$$

$$\mathbf{Z}'_{m,t} = [\sin(m\lambda_t), \cos(m\lambda_t)], \quad (13d)$$

$$\lambda_t = 2\pi t_s / T_s, \quad (13e)$$

and $m \in \{1, \dots, M\}$. This model (henceforth denoted M3a) can also include power terms involving t_s to avoid the assumption that γ_t takes the same value at the opening and close of the trading day. Indeed, such an FFF model (including t_s and t_s^2) has been proposed by Andersen and Bollerslev (1997, 1998) when modeling intraday return volatility dynamics. Following their lead, M3a also includes t_s and t_s^2 in its specification.

Though the FFF-version of the model is parsimonious and allows for smooth γ_t dynamics, it is somewhat rigid in functional form and, therefore, may not be able to capture complex dynamics. To overcome this potential shortcoming, a spline-version of the model is also considered that allows different cubic spline functions to be estimated between selected points (referred to as knots) within the periodic cycle. In particular, letting k_n denote the (equally spaced) n th knot, with $k_n = \{k_n \in \mathbb{Z}^+ : 1 \leq k_n \leq T_s\}$, $n \in \{1, \dots, N\}$, and $k_1 = 0$, the spline-version of the model

Franses, de Bruin and van Dijk (2000) allow the linear parameters of the STAR model to vary over the periodic cycle via the use of such dummies. However, economic theory predicts that it is the smoothness parameter, and not the linear parameters, that will vary over time. As such, this model is not considered in this paper.

(henceforth denoted M3b) is given by (11) with

$$\boldsymbol{\theta}' = [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_n, \dots, \boldsymbol{\theta}'_N], \quad (14a)$$

$$\boldsymbol{\theta}'_n = [\theta_{1,n}, \theta_{2,n}, \theta_{3,n}], \quad (14b)$$

$$\mathbf{Z}'_t = [\mathbf{Z}'_{1,t}, \dots, \mathbf{Z}'_{n,t}, \dots, \mathbf{Z}'_{N,t}], \quad (14c)$$

$$\mathbf{Z}'_{n,t} = D_n[(t_s - k_n), (t_s - k_n)^2, (t_s - k_n)^3], \quad (14d)$$

and D_n equals unity if $t_s \geq k_n$, and zero otherwise. Though no studies have used this functional form in the context of the STAR model, it has been used extensively in trade duration models. Most notably, cubic splines have been successfully incorporated into a variety of autoregressive duration (ACD) models (see, e.g., Engle and Russell, 1995, 1998, Zhang, Russell and Tsay, 2001, and Taylor, 2003).

4 A new linearity test

This section contains the derivation of an LM test for linearity, against the alternative of non-linearity based on the augmented STAR model introduced in this paper. In addition, a series of Monte Carlo experiments are described that demonstrate that this test is correctly sized.

4.1 Derivation

Testing the null hypothesis of linearity, against the alternative of STAR non-linearity, suffers from the problem of unidentified nuisance parameters. This means that certain parameters are not restricted to equal zero under the null, even though they appear in the non-linear part of the model. In the current study, if the null of linearity holds with $\varphi_1 = \varphi_2$, then γ_t does not necessarily have to equal zero. To overcome this problem, we use a methodology similar to that used by Luukkonen, Saikkonen and Teräsvirta (1988). In particular, we derive a LM test statistic (with an asymptotic χ^2 distribution), designed to test the null hypothesis of linearity,

against the alternative of the augmented STAR non-linearity introduced in this paper.

The derivation of the LM test statistic proceeds in three stages, culminating in an auxiliary regression equation that can be restricted under the null without incurring the problem of unidentified nuisance parameters. The derivation begins by noting that the augmented STAR model given by (9) can be rewritten as follows:

$$y_t = \boldsymbol{\varphi}_1' \mathbf{x}_t + \boldsymbol{\varphi}_2' \mathbf{x}_t G(y_{t-d}; \gamma_t, c) + \epsilon_t. \quad (15)$$

where $\boldsymbol{\varphi}_2^* = \boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_j = [\varphi_{0,j}, \varphi_{1,j}, \dots, \varphi_{p,j}]'$, $j \in \{1, 2\}$, $\mathbf{x}_t = [1, \tilde{\mathbf{x}}_t']'$, $\tilde{\mathbf{x}}_t = [y_{t-1}, \dots, y_{t-p}]'$, and $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$, $t \in \{1, \dots, T\}$, and $G(y_{t-d}; \gamma_t, c)$ is given by (10). The next stage involves approximating the transition function, $G(y_{t-d}; \gamma_t, c)$, by a first-order Taylor expansion around $\gamma_t = 0$, that is,

$$\begin{aligned} G(y_{t-d}; \gamma_t, c) &\approx T(y_{t-d}; \gamma_t, c) = G(y_{t-d}; 0, c) + \gamma_t \left. \frac{\partial G(y_{t-d}; \gamma_t, c)}{\partial \gamma_t} \right|_{\gamma_t=0} + R(y_{t-d}; \gamma_t, c), \\ &= \gamma_t (y_{t-d} - c)^2 + R(y_{t-d}; \gamma_t, c), \end{aligned} \quad (16)$$

where $R(y_{t-d}; \gamma_t, c)$ is a remainder term, which equals zero under the null hypothesis that $\gamma_t = 0$.¹¹ The final stage is achieved by substituting (16) into (15), and using the expression for γ_t given by (11), to give the auxiliary regression equation,

$$\begin{aligned} y_t &= \boldsymbol{\alpha}'_1 \mathbf{x}_t + \boldsymbol{\alpha}'_2 \mathbf{x}_t y_{t-d} + \boldsymbol{\alpha}'_3 \mathbf{x}_t y_{t-d}^2 + \boldsymbol{\beta}'_1 (\mathbf{Z}_{t-d} \otimes \mathbf{x}_t) \\ &\quad + \boldsymbol{\beta}'_2 (\mathbf{Z}_{t-d} \otimes \mathbf{x}_t y_{t-d}) + \boldsymbol{\beta}'_3 (\mathbf{Z}_{t-d} \otimes \mathbf{x}_t y_{t-d}^2) + e_t, \end{aligned} \quad (17)$$

where $e_t = \epsilon_t + (\boldsymbol{\varphi}'_2 - \boldsymbol{\varphi}'_1) \mathbf{x}_t R(y_{t-d}; \gamma_t, c)$, $\boldsymbol{\alpha}_k = (\alpha_{0,k}, \dots, \alpha_{p,k})'$, $\boldsymbol{\beta}_k = \boldsymbol{\theta} \otimes \tilde{\boldsymbol{\beta}}_k$, $\tilde{\boldsymbol{\beta}}_k = (\tilde{\beta}_{0,k}, \dots, \tilde{\beta}_{p,k})'$,

¹¹The transition function could be approximated by a higher order Taylor expansion to produce a more accurate representation of this function. Indeed, Escribano and Jordá (1999) use a second-order Taylor expansion in the context of a conventional STAR model. However, the first-order expansion is deemed appropriate given the results presented in the following section. Most notably, the test is correctly sized and has power to reject a false null.

$k \in \{1, \dots, 3\}$,

$$\alpha_{l,1} = \varphi_{l,1} + \theta_0 c^2 (\varphi_{l,2} - \varphi_{l,1}), \quad (18a)$$

$$\alpha_{l,2} = -2\theta_0 c (\varphi_{l,2} - \varphi_{l,1}), \quad (18b)$$

$$\alpha_{l,3} = \theta_0 (\varphi_{l,2} - \varphi_{l,1}), \quad (18c)$$

$$\tilde{\beta}_{l,1} = c^2 (\varphi_{l,2} - \varphi_{l,1}), \quad (18d)$$

$$\tilde{\beta}_{l,2} = -2c (\varphi_{l,2} - \varphi_{l,1}), \quad (18e)$$

$$\tilde{\beta}_{l,3} = (\varphi_{l,2} - \varphi_{l,1}), \quad (18f)$$

and $l \in \{0, \dots, p\}$.

Tests of the null hypothesis of linearity, against the alternative of augmented STAR non-linearity, can be performed by testing restrictions on the above auxiliary regression equation. To achieve a better understanding of the required restrictions, it is worth noting that the null hypothesis, $H_0 : \gamma_t = 0$, is equivalent to the null, $H'_0 : \theta_0 = 0, \boldsymbol{\theta} = \mathbf{0}$. In turn, one can see from (18a) to (18f), that testing this hypothesis is equivalent to testing the null, $H''_0 : \alpha_{l,2} = \alpha_{l,3} = 0, \tilde{\beta}_k = \mathbf{0}$. (From this version of the null, one can see that the unidentified nuisance parameter problem is overcome as the restriction under the null can also be achieved by imposing the restriction that $\varphi_1 = \varphi_2$). Given the nature of the testing procedure it follows that a standard LM test statistic (with χ^2 distribution) can be used to test this null hypothesis.

One can also use the above regression equation to test the null hypothesis of linearity, against the alternative of (non-augmented) STAR non-linearity, as this is equivalent to testing the restriction that $\theta_0 = 0$, conditional on smoothness not being functionally dependent on any of the explanatory variables, that is, $H_0 : \theta_0 = 0 | \boldsymbol{\theta} = \mathbf{0}$. Inspection of (18a) to (18f) implies that this null hypothesis is equivalent to the null, $H'_0 : \alpha_{l,2} = \alpha_{l,3} = 0 | \tilde{\beta}_k = \mathbf{0}$. As with the previous test, a standard LM test statistic (with χ^2 distribution) can be used to test this null hypothesis.

4.2 Test properties

A crucial feature of any statistical test is that it rejects the true null hypothesis the correct number of times, i.e., with probability given by the nominal size of the test. To examine this feature, the empirical sizes of the above tests are calculated via the use of two sets of Monte Carlo experiments (each consisting of 10,000 replications of the data). In the first set of experiments, $\text{IN}(0,1)$ errors are used to generate data via (15) under the assumption that $\varphi_1 = \varphi_2 = 0$, with $T \in \{125, 250, 500, 1000, 10000\}$, and with FFF and cubic spline periodic components generated using $M = 3$, $N = 3$, and $T_s = T/5$. The second experiment repeats the first experiment, but uses randomized (bootstrapped) residuals from the FFF-version of (15) obtained in the subsequent empirical section. In both cases, the tests for linearity are based on auxiliary regression equations with $p = 2$ and $d = 1$.¹²

The empirical sizes (at 1%, 5%, and 10% nominal size) of various LM tests for linearity are calculated in the above experiments. First, the empirical sizes of the tests for linearity, against the FFF-version, and the spline-version of the augmented STAR model (henceforth denoted LM_{3a} and LM_{3b} , as they are based on the M3a STAR model and M3b STAR model, respectively), are calculated. Second, the latter of the tests proposed in the previous subsection is assessed by considering the empirical size of the test of the null that $\theta_0 = 0$, with the assumption $\theta = \mathbf{0}$ imposed (without testing) on the regression equation. (As this test is equivalent to testing the null of linearity against the M2 STAR model, it is henceforth denoted LM_2). In addition, the empirical sizes of the tests of the null that $\theta = \mathbf{0}$, using the FFF-version (henceforth denoted $\text{LM}_{3a,1}$), and the spline-version (henceforth denoted $\text{LM}_{3b,1}$) of the augmented STAR model, are calculated.

The results given in Table 1 indicate that, in general, all the tests appear to be reasonably well sized. However, there is some evidence of size distortion when small sample sizes are used. As sample sizes in finance typically exceed those used in these experiments, such size distortion should be of minor concern from a practitioner point of view. When large sample sizes are

¹²The parameter values used in these experiments are based on those used in the subsequent empirical section of this paper.

not available, a small-sample correction for the distortion is required. One possible method of correction is achieved by performing F-distribution versions of the tests (denoted by LM_2^* , LM_{3a}^* , LM_{3b}^* , $LM_{3a,1}^*$, and $LM_{3b,1}^*$). The results show that these corrections are broadly successful.

5 Empirical results

This section contains results pertaining to the application of the models and tests described in the previous two sections. After the data used have been described, details of the specifications of the models estimated, tests for linearity (against various non-linear alternatives), and tests of model adequacy, are provided.

5.1 Data

We make use of various pieces of information concerning all trades in various securities relating to the S&P 500 index, covering the period, January 2, 2001 to December 31, 2002. In particular, returns, defined as the log change in transaction prices; and trading volume, defined as the number of contracts traded within an interval, were obtained for regular S&P 500 futures contracts (with ticker symbol SP), and S&P 500 E-mini futures contracts (with ticker symbol ES), both of which trade on the Chicago Mercantile Exchange. Both of these S&P 500 futures contracts specify cash settlement of the contract at 8:30 a.m. on the 3rd Friday of March, June, September, and December. To obtain a single continuous series for each type of futures contract, we assume that futures contracts with the nearest maturity are replaced (through trading), by contracts with the next nearest maturity – a practice referred to as rolling over – when the next contract’s daily tick count exceeds the current contract tick count. These data were obtained from *Tick Data Incorporated*.

In addition to the above futures contract data, data were also obtained for the S&P 500 iShares ETF (with ticker symbol IVV), which (mainly) trades on the American Stock Exchange. Use of these data represents an innovation in the study of mispricing dynamics, as previous studies construct mispricing using the level of the S&P 500 index. In using ETF data, we are

also explicitly assuming that arbitrageurs monitor and use the price of this fund to correct any potential mispricing. This seems reasonable given the obvious cost advantages of trading this fund, in comparison to trading each of the stocks making up the fund.¹³ Also, choice of this particular fund over competing funds (e.g., the Spider ETF) is based on the fact that the iShares ETF allows the timely re-investment of dividend income – an important feature given that the definition of the fair value of the futures contract given by (1) assumes that such timely re-investment occurs.¹⁴ These data were also obtained from *Tick Data Incorporated*.

The above transaction data are converted to five-minute frequency data.¹⁵ This frequency is deemed to be sufficiently low enough to avoid stale data, and high enough to avoid loss of information. As the trading hours of each security differ, these data are necessarily truncated to produce a fully synchronized intraday dataset. In particular, the trading hours for regular S&P 500 futures contracts, E-mini S&P 500 futures contracts, and S&P 500 iShares ETF shares are 8:30 a.m. to 3:15 p.m., 12:00 (midnight) to 3:15 p.m., and 9:30 a.m. to 4:15 p.m., respectively. Therefore, to generate an overlapping dataset, these futures and spot series are truncated to contain data over the intraday period, 9:35 a.m. to 3:15 p.m. The available data is further truncated by the application of the procedure proposed by Miller, Muthuswamy and Whaley (1994) to remove the bid-ask bounce in the asset prices. Specifically, they suggest fitting an MA(1) model to asset returns and then using the residuals from this model as the asset returns net of bid-ask bounce. Upon completion of this procedure, spot and futures data are available over the intraday period, 9:40 a.m. to 3:15 p.m.¹⁶

The (two) futures series and the (one) spot series are then used to construct two mispricing series via (1). One possible method of constructing the theoretical futures price is to use a government-issued interest rate with a maturity closest to that of the futures contract, and

¹³Use of ETF share prices also mitigates a major market microstructure effect highlighted by Miller, Muthuswamy and Whaley (1994). Specifically, the problem of infrequent trading in stocks within the index is removed.

¹⁴The Spider ETF only allows dividend re-investment at the end of each quarter.

¹⁵The subsequent analysis is virtually unchanged when fifteen-minute frequency data are used. These results are available upon request.

¹⁶This procedure is applied to the spot and futures return series considered in this paper. To avoid picking up dependence spanning the overnight period, the dependent variable in this (first-order) model is restricted to be that observed from 9:40 a.m. (each day) onwards. After fitting the model, the new spot and futures price series are obtained by means of daily numerical integration of the model residuals covering the intraday period, 9:40 a.m. to 3:15 p.m.

expected (or realized) dividend payments for each stock in the index over the life of the futures contract (see, e.g., Tse, 2001, and Taylor, 2003). However, as these variables can only change on a daily basis, we use the daily demeaned difference between the (log) spot and futures prices as the measure of mispricing. The resulting mispricing series are henceforth referred to as regular futures mispricing and E-mini futures mispricing.¹⁷

5.2 Evidence of intraday periodicity

To give an indication of the likely intraday periodicities within the data, Figure 1 shows the intraday periodicity of the absolute mispricing and the trading volume associated with regular futures mispricing (Panel A and Panel C, respectively) and E-mini futures mispricing (Panel B and Panel D, respectively).¹⁸ These periodicities are based on the intraday means of these series for each five-minute period within the trading day. The graphs show a clear pattern: absolute mispricing and trading volume follows a U-shaped pattern – a commonly observed pattern in financial markets.¹⁹ However, the extent to which one can infer patterns in the level of arbitrage activity is limited. The (time-invariant) STAR model implies that this level will also follow a U-shaped pattern. This is because this model assumes that arbitrage activity is a positive (time-invariant) function of absolute mispricing. By contrast, the augmented STAR model (introduced in this paper) allows time-variation in the functional dependence between arbitrage activity and mispricing. As such, the level of arbitrage activity does not necessarily have to follow a U-shaped pattern. Indeed, the results in the subsequent analysis demonstrate that this potential for alternative patterns in the level of arbitrage activity is achieved.

5.3 Linearity testing

To give some indication of the likely suitability of the various econometric models of mispricing, tests for linearity, against the non-linearity implied by these models, are conducted. In addition to the tests described in the previous section (LM_2 , LM_{3a} , LM_{3b} , $LM_{3a,1}$, and $LM_{3b,1}$),

¹⁷All subsequent analysis is based on these mispricing being multiplied by 100.

¹⁸Trading volume is defined as sum of the number of futures contracts and iShares ETF shares traded.

¹⁹Trading volume drops dramatically at 3:00 p.m. when the underlying S&P 500 market closes.

we also test the null hypothesis of linearity against the alternative of M1-type non-linearity using the likelihood ratio test (henceforth denoted LR_1) proposed by Hansen (1996, 1997, 1999). In each case, the tests are applied to both mispricing series with $p \in \{1, 2, 3\}$, $d \leq p$, $M = 3$, and $N = 3$, where use of $d \leq p$ implies that trade delays are not expected to exceed the degree of time-dependence in mispricing.²⁰

The results in Table 2 indicate that, in general, the null hypothesis of linearity can be rejected. However, the degree of rejection is very much a function of the type of test performed. In particular, the LR_1 and LM_2 tests indicate that the null can only be rejected (at the 1% level) when $p = 3$ and $d > 2$. However, when the new linearity tests are performed they indicate rejection of the null (even at the 0.01% level) for all values of p and d . This result suggests that allowing the propensity-to-arbitrage, as measured by γ_t , to time-vary over the trading day is an accurate representation of the arbitrage process.

5.4 Model estimates

Five models are estimated in this paper. In addition to the non-linear models M1, M2, M3a, and M3b, described previously, we also estimate a linear AR model (henceforth denoted M0). These models are estimated using a variety of methods: the AR model (M0) is estimated by ordinary least squares (OLS), the TAR model (M1) is estimated using the methodologies proposed by Hansen (1996, 1997, 1999, 2000), and the STAR model (M2) is estimated by non-linear least squares (NLS).²¹ The augmented STAR models (M3a and M3b) are also estimated by NLS, in a manner similar to that used to estimate M2. In particular, the parameter set, $\boldsymbol{\pi} = (\boldsymbol{\varphi}'_1, \boldsymbol{\varphi}'_2, c, \theta_0, \boldsymbol{\theta}')$, is estimated follows:

$$\hat{\boldsymbol{\pi}} = \arg \min_{\boldsymbol{\pi}} Q_T(\boldsymbol{\pi}) = \arg \min_{\boldsymbol{\pi}} \sum_{t=1}^T (y_t - H(\mathbf{x}_t; \boldsymbol{\pi}))^2, \quad (19)$$

²⁰To avoid picking up dependence spanning the overnight period, the dependent variable (y_t) in these tests is restricted to be observed from 9:55 a.m. (each day) onwards. Use of this daily starting point ensures that an equal number of observations are included in each test, regardless of the value of p .

²¹See Franses and van Dijk (2000) for an comprehensive description of the estimation of STAR models.

where

$$H(\mathbf{x}_t; \boldsymbol{\pi}) = \boldsymbol{\varphi}'_1 \mathbf{x}_t (1 - G(y_{t-d}; \gamma_t, c)) + \boldsymbol{\varphi}'_2 \mathbf{x}_t G(y_{t-d}; \gamma_t, c). \quad (20)$$

As with the estimation of time-invariant STAR models, the estimation process can be greatly simplified by concentrating the sum of squares function, $Q_T(\boldsymbol{\pi})$, with respect to $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$. Specifically, for given values of c , θ_0 , and $\boldsymbol{\theta}$, the parameters $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ can be estimated by OLS. Consequently, the minimization process can focus exclusively on the non-linear parameters of the model as the linear parameters can be iteratively estimated by OLS.²² Finally, statistical inference relating to $\boldsymbol{\pi}$ can be carried out by noting that under certain regularity conditions (see White and Domowitz, 1984),

$$\sqrt{T}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \overset{a}{\approx} \text{N}(0, \boldsymbol{\Sigma}), \quad (21)$$

where $\boldsymbol{\pi}_0$ is the true value of the parameter set, and $\boldsymbol{\Sigma}$ is the asymptotic covariance-matrix of $\hat{\boldsymbol{\pi}}$. The inference used in this paper is based on the heteroscedasticity-consistent (HCC) covariance-matrix $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$, where

$$\hat{\mathbf{A}} = \frac{1}{T} \sum_{t=1}^T \nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}}) \nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}})', \quad (22)$$

$$\hat{\mathbf{B}} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 \nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}}) \nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}})', \quad (23)$$

and $\nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}}) = \partial H(\mathbf{x}_t; \hat{\boldsymbol{\pi}}) / \partial \boldsymbol{\pi}$.²³

The models are estimated using the two mispricing series for $(p, d, M, N) \in \{1, 2, 3\}$.²⁴ It is found that the optimal fit (according the Akaike information criterion, henceforth denoted AIC) for all models is obtained when $p = 2$, $d = 1$, $M = 3$, and $N = 3$.²⁵ Results pertaining to these models are presented in Table 3, and contain the estimated linear parameter values (and

²²The estimation of the STAR-based models (M2, M3a, and M3b) is further simplified by imposing the restriction that $c = 0$. This restriction is motivated by the observation that zero mispricing should not, according to the economic arguments provided in Section 2, trigger any arbitrage activity.

²³Details of the specifications for the gradients, $\nabla H(\mathbf{x}_t; \hat{\boldsymbol{\pi}})$, are available upon request.

²⁴As in the case of the linearity tests, to avoid picking up dependence spanning the overnight period, the dependent variable in the models is restricted to be that observed from 9:55 a.m. (each day) onwards. Similarly, use of this daily starting point ensures that an equal number of observations are included in each model, regardless of the value of p .

²⁵Results pertaining to the other values of p , d , M , and N , are available upon request.

associated HCC standard errors), summary details of the non-linear parameter values (and associated HCC standard errors where applicable), and various metrics pertaining to model fit.²⁶ The estimated AR model (M0) indicates that mispricing is highly time-dependent. However, the degree of time-dependence is such that it does not appear to follow a unit-root process, thus indicating the possible presence of arbitrageurs. This presence is confirmed by the estimated TAR model (M1), which shows that mispricing appears to follow two different processes depending on whether a profit opportunity exists, with a lower degree of time-dependence in mispricing when such an opportunity exists.²⁷ Rather surprisingly, allowing the arbitrage bound to vary over arbitrageurs, hence use of the STAR model (M2), does not lead to an improvement in the degree of model fit (as measured by the AIC). By contrast, allowing time-variation in the propensity-to-arbitrage, as implied by the FFF-version and the spline-version of the augmented STAR model (M3a and M3b, respectively), leads to a large improvement in model fit. It is also noticeable that, in comparison to M2, the degree of time-dependence appears more polarized between the two regimes, with mispricing following a near unit root process when no profit opportunity exists for any arbitrageurs ($G(y_{t-d}; \gamma_t, c) = 0$), and a less time-dependence process when profit opportunities exist for all arbitrageurs ($G(y_{t-d}; \gamma_t, c) = 1$).

The above results are broadly consistent over the two measures of mispricing, except that there is some evidence of a greater level of arbitrage activity when E-mini futures mispricing is used. This is evinced by the smaller value of the estimated arbitrage bound, $\hat{\kappa}$, and the larger values of average smoothness, $\overline{\hat{\gamma}_t}$, and the average value of the transition function, $\overline{\hat{G}(\cdot)}$, when this mispricing measure is used. This result is consistent with the findings of Hasbrouck (2003), who finds that the E-mini market more rapidly reflects information pertaining to stocks in the S&P 500 index.

²⁶Space limitations prohibit presentation of the estimated values of θ_0 and θ for the FFF-version and the spline-version of the augmented STAR model. However, these estimates (and their associated HCC standard errors) are available upon request.

²⁷Short-selling restrictions on stocks within the underlying index imply use of a TAR model with an asymmetric arbitrage bound. The appropriateness of such a model is examined by estimating a three-regime TAR model, where the regimes are not restricted to have equal (absolute) values. The results indicate that such a model provides an inferior fit (as given by the AIC), in comparison to the symmetric TAR model (M1) considered in this paper. (These results are available upon request). These findings are compatible with those obtained by Neal (1996), who takes a direct investigative approach to arbitrageur behavior. Using data that actually identifies arbitrage trades, Neal finds that, inter alia, short selling restrictions ‘... are unlikely to affect the cash-futures mispricing’.

To give some idea of the nature of the intraday periodicity in arbitrageur behavior, Figure 2 plots the estimated smoothness (propensity-to-arbitrage) and the associated mean transition function values (the level of arbitrage activity) over the trading day, obtained using regular futures mispricing (Panel A and Panel C, respectively), and E-mini futures mispricing (Panel B and Panel D, respectively). When using M2, the propensity-to-arbitrage is, by definition, constant over the trading day. However, when M3a and M3b are used, this propensity takes on a clear intraday periodic pattern: high at the opening of trading, and low at the close of trading, with a local minimum in this propensity occurring during the lunch-time period. Despite this clear evidence, it cannot tell us anything about the level of arbitrage unless we combine this propensity with the trigger for arbitrage (the absolute level of mispricing). This is achieved by considering the mean values of the transition functions over the trading day. When using M2, the level of arbitrage activity follows a U-shaped pattern. This is the result of a constant propensity-to-arbitrage and the U-shaped pattern of absolute mispricing observed in Figure 1. By contrast, the level of arbitrage is not necessarily U-shaped when using M3a and M3b. In this case, the time-variation in the propensity-to-arbitrage and absolute mispricing combine to form a pattern similar to that observed for the propensity-to-arbitrage pattern. In particular, arbitrage activity is high around the opening of trading, and low during the lunch-time period and around the close of trading – a result that may indicate an aversion to holding overnight arbitrage positions. The figure also shows that there are differences in the pattern over the measures of mispricing. In particular, the propensity-to-arbitrage and the level of arbitrage in the E-mini futures market appears greater than in the regular futures market. Again, this result may reflect a preference for index arbitrage trading in the former market due its greater informational efficiency.

5.5 The impact of shocks

The impact of shocks on the S&P 500 index arbitrage market is assessed by considering the values of the model-based generalized impulse response function (GIRF). This function is

defined as follows:

$$\text{GIRF}_y(h, \delta, \omega_{t-1}) = \text{E}[y_{t+h} | \epsilon_t = \delta, \omega_{t-1}] - \text{E}[y_{t+h} | \omega_{t-1}], \quad (24)$$

where ω_{t-1} measures the history of all previous shocks, δ is the initial magnitude of the shock, and $h \in \{0, 1, 2, \dots, H\}$.²⁸ This function gives the impact of shocks to mispricing using a subset of the history of past shocks. In the current application, the impact of the shocks is measured at 1,000 randomly assigned points in the history, with the whole process repeated 1,000 times. To obtain the GIRF, the means of these impacts, net of the mean mispricing over the corresponding points, are calculated using the estimated coefficients from each model. The resulting function shows the impact of a shock to mispricing, irrespective of the point in time at which the shock occurs.²⁹ The asymmetry of the shocks can also be assessed by calculating the values of the following function:

$$\text{ASY}_y(h, \delta, \omega_{t-1}) = \text{GIRF}_y(h, \delta, \omega_{t-1}) + \text{GIRF}_y(h, -\delta, \omega_{t-1}). \quad (25)$$

Diagrammatic representations of the functions, given by (24) and (25), are presented in Figure 3 (regular futures mispricing) and Figure 4 (E-mini futures mispricing), as implied by M0, M2, M3a, and M3b.³⁰ These figures show the impact of a unit shock (where $\sigma_{\hat{\epsilon}} \approx 0.43$) conditional on all shock histories (Panel A), conditional on shocks occurring during the first fifteen minutes of the trading day (Panel B), and conditional on shocks occurring during the last fifteen minutes of the trading day (Panel C). The results indicate that all models imply that a unit shock disappears after 3 hours (when $h = 36$). However, the impacts of the shock, as implied by the models, differ over the course of the trading day. Most notably, M3a and M3b imply that shocks disappear more rapidly if they occur during the opening period. This

²⁸See Koop, Pesaran and Potter (1996) and Franses and van Dijk (2000) for details of impulse response analysis using non-linear models.

²⁹Note that non-linear models do not have a Wold representation. As such, shocks in different periods *do* interact with each other – hence the need for a simulation approach.

³⁰M1 gives virtually identical results to M2 and is omitted on the grounds that it makes the diagrams easier to interpret. Results pertaining to this model are available upon request.

is because there is a greater desire to trade the shock at this time. The reverse holds at the end of the trading day, where shocks are less rapidly incorporated into the market. By contrast, the time-invariant models (M0 and M2) cannot pick up such variation in the impact of shocks over the trading day. The figures also show that there is a degree of asymmetry in the S&P 500 index arbitrage market (Panel D). Specifically, the results indicate that, when M3a and M3b are used, positive shocks are more rapidly incorporated into the market than negative shocks, though the difference is very small.³¹

5.6 Forecasting ability

Given the importance of accurate forecasts of mispricing to arbitrageurs and hedgers, it seems reasonable to assess model performance via a comparison of forecasting ability.³² To this end, the five models (M0, M1, M2, M3a, and M3b) are re-estimated using the first half of the dataset, both measures of mispricing, and $(p, d, M, N) \in \{1, 2, 3\}$.³³ The estimated coefficients of the optimal-fit models (based on the Schwarz information criterion) and the second half of the dataset are then used to generate time-consistent 1-step ahead forecasts of mispricing. These forecasts are then compared to realized mispricing observed in the second half of the sample.

To formally test the comparative accuracy of the model-based forecasts, we make use of the extant tests of Granger and Newbold (1977), henceforth denoted the MGN test, and Meese and Rogoff (1988), henceforth denoted the MR test, and the asymptotic test introduced by Diebold and Mariano (1995), henceforth denoted the DM test. The latter test is preferable to the extant tests as it is, inter alia, robust to non-zero mean forecast errors, non-normally distributed forecast errors, and serially correlated forecast errors. In the current application, it is the robustness of the DM test statistic to the non-normality assumption that is most attractive. Indeed, when

³¹This result may be due to it being more expensive to short-sell stocks in the S&P 500 index than it is to sell futures contracts. However, as described in footnote #27, no evidence of this was obtained when a TAR model with asymmetric arbitrage bounds was estimated.

³²In conducting such a comparison, we are explicitly conceding the fact that the ‘best’ model may still be ‘inadequate’ in terms of model misspecification. However, as a wide range of currently available non-linear models are used, the approach will give some indication as to which model most accurately represents the data.

³³Details of the estimated models are available upon request. Other proportions of the dataset are also considered. For instance, we use the first quarter and the first three quarters of the dataset in the estimation section of the exercise. The results obtained using such sample periods produce similar results to those presented in this paper. These results are also available upon request.

the forecast errors are tested for normality, the null is rejected at the 1% significance level on every occasion.³⁴

A summary of the results obtained when the MGN, MR, and DM tests are performed on all available forecasts is given in Table 4. We report the test statistics and the associated p-values, where each model-based set of forecasts are compared with each other. The results indicate that M3a and M3b each successfully beat the other models, for all types of test and for both measures of mispricing, though the tests cannot distinguish between the dominance of the M3a and M3b models. Given this evidence, it would appear that the models introduced in this paper produce significantly more accurate forecasts than existing models. As such, they should be used by those seeking to generate accurate forecasts of mispricing, and would appear to be of most value to users of financial markets.

6 Concluding remarks

This paper provides an accurate econometric model of (contemporaneous) mispricing in spot and futures markets. This model differentiates itself from previous models in its treatment of time in the arbitrage process. Existing models of mispricing assume that the propensity-to-arbitrage (for a given level of mispricing) is constant. However, the economic model that motivates arbitrage activity suggests that this propensity is determined by exogenous factors. As these factors are likely to exhibit strong intraday periodicities, it is likely that the propensity-to-arbitrage will also vary over the trading day. This is indeed the case when a model that allows such behavior is fitted to mispricing in the S&P 500 spot and futures market. However, the nature of the periodicity does not follow the usual U-shaped pattern. Rather, it tends to be the case that the propensity-to-arbitrage is high at the opening of trading and low at the close of trading, even though the degree of mispricing is high during both of these periods. Thus it would appear that arbitrageurs are willing to trade at the beginning, and not at the end, of the trading day – implying that impediments to arbitrage are insufficiently large at the beginning

³⁴These results are available upon request.

of the trading day to prevent arbitrage activity occurring, but are prohibitively large at the end of the trading day. There is also the possibility that arbitrageurs take positions early in the trading day in the hope that these are profitably unwound during the same trading day. This would be the case if it could be shown that arbitrageurs are unwilling to trade at the end of the trading day, even though the impediments to trade are not prohibitively large. This issue is left for future research.

References

- Andersen, T., and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.
- , 1998, DM-Dollar volatility: Intraday activity patterns, macroeconomic announcements and longer run dependencies, *Journal of Finance* 53, 219–265.
- Bailey, W., The market for Japanese stock index futures: Some preliminary evidence, *Journal of Futures Markets* 9, 283–295.
- Baillie, R., and T. Bollerslev, 1991, Intraday and intermarket volatility in foreign exchange rates, *Review of Economic and Statistics* 58, 565–585.
- Barberis, N., and R. Thaler, 2003, A survey of behavioral finance, in G. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of Economics and Finance*, North-Holland, Amsterdam.
- Brennan, M., and E. Schwartz, 1990, Arbitrage in stock index futures, *Journal of Business* 63, S7–S31.
- D’Avolio, G., 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.
- De Long, J., Shleifer, A., Summers, L., and R. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.
- Diebold, F., and R. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–265.
- Dwyer, G., Locke, P., and W. Yu, 1996, Index arbitrage and non-linear dynamics between the S&P 500 futures and cash, *Review of Financial Studies* 9, 301–332.
- Engle, R., and J. Russell, 1995, Autoregressive conditional duration: A new model for irregularly spaced transaction data, UCSD Working Paper.
- , 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127–1162.
- Escribano, A., and O. Jordá, 1999, Improved testing and specification of smooth transition regression models, in P. Rothman (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Kluwer Academic Press, Boston.
- Forbes, C., Kalb, R., and P. Kofman, 1999, Bayesian arbitrage threshold models, *Journal of Business and Economic Statistics* 17, 364–372.
- Franses, P. H., de Bruin, P., and D. van Dijk, 2000, Seasonal smooth transition autoregression, Econometric Institute Report 2000-06/A.
- Franses, P. H., and D. van Dijk, 2000, *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.
- Fung, J., and P. Draper, 1999, Index arbitrage opportunities and short sales constraints, *Journal of Futures Markets* 10, 695–715.

- Garrett, I., and N. Taylor, 2001, Intraday and interday basis dynamics: Evidence from the FTSE 100 index futures market, *Studies in Nonlinear Dynamics and Econometrics* 5, 133–152.
- Granger, C., and P. Newbold, 1977, *Forecasting Economic Time Series*, Academic Press, Orlando.
- Granger, C., and T. Teräsvirta, 1993, *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Hansen, B., 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.
- , 1997, Inference in TAR models, *Studies in Non-linear Dynamics and Econometrics* 2, 1–14.
- , 1999, Testing for linearity, *Journal of Economic Surveys* 13, 551–576.
- , 2000, Sample splitting and threshold estimation, *Econometrica* 68, 575–603.
- Harris, L., 1989, The October 1987 S&P 500 stock-futures basis, *Journal of Finance* 44, 77–100.
- Harvey, C., and R. Huang, 1991, Volatility in the foreign currency futures market, *Review of Financial Studies* 4, 543–569.
- Hasbrouck, J., 2003, Intraday price formation in US equity index markets, *Journal of Finance* 58, 2375–2400.
- Kleidon, A., 1992, Arbitrage, non-trading, and stale prices: October 1987, *Journal of Business* 65, 483–508.
- Koop, G., Pesaran, H., and S. Potter, 1996, Impulse response analysis in non-linear multivariate models, *Journal of Econometrics* 74, 119–147.
- Lim, K., 1990, Arbitrage and price behavior of Nikkei stock index futures, *Journal of Futures Markets* 12, 151–161.
- Lundbergh, S., Teräsvirta, T., and D. van Dijk, 2003, Time-varying smooth transition autoregressive models, *Journal of Business and Economic Statistics* 21, 104–121.
- Luukkonen, R., Saikkonen, P., and T. Teräsvirta, 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* 75, 491–499.
- MacKinlay, A., and K. Ramaswamy, 1988, Index-futures arbitrage and the behavior of stock index futures prices, *Review of Financial Studies* 1, 137–158.
- Martens, M., Kofman, P., and T. Vorst, 1998, A threshold error-correction model for intraday futures and index returns, *Journal of Applied Econometrics* 13, 245–263.
- McInish, T., and R. Wood, 1992, An analysis of intraday patterns in bids, asks, and spreads for NYSE common stocks, *Journal of Finance* 47, 753–764.
- Meese, R., and K. Rogoff, 1988, Was it real? The exchange rate-interest differential relation over the modern floating-rate period, *Journal of Finance* 43, 933–948.

- Miller, M., Muthuswamy, J., and R. Whaley, 1994, Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance* 49, 479–513.
- Neal, R., 1996, Direct tests of index arbitrage models, *Journal of Financial and Quantitative Analysis* 31, 541–562.
- Shleifer, A., and L. Vishny, 1997, The limits to arbitrage, *Journal of Finance* 52, 32–55.
- Sofianos, N., 1993, Index arbitrage profitability, *Journal of Derivatives* 1, 6–20.
- Taylor, N., 2003, Trading intensity, volatility, and arbitrage activity, *Journal of Banking and Finance*, forthcoming.
- Taylor, N., van Dijk, D., Franses, P., and A. Lucas, 2000, SETS, arbitrage activity, and stock price dynamics, *Journal of Banking and Finance* 24, 1289–1306.
- Teräsvirta, T., 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* 89, 208–218.
- , 1998, Modeling economic relationships with smooth transition regressions, in A. Ullah, and D. Giles (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York.
- Tong, H., 1983, *Threshold Models in Nonlinear Time Series Analysis*, Springer-Verlag, New York.
- , 1990, *Nonlinear Time Series: A Dynamical Systems Approach*, Oxford University Press, Oxford.
- Tong, H., and K.S. Lim, 1980, Threshold autoregression, limit cycles and cyclical data, *Journal of the Royal Statistical Society B* 42, 245–292.
- Tsay, R., 1989, Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* 84, 231–240.
- Tse, Y., 1999, Market microstructure of the FTSE 100 index futures: An intraday empirical analysis, *Journal of Futures Markets* 19, 31–58.
- , 2001, Index arbitrage with heterogeneous investors: A smooth transition error correction analysis, *Journal of Banking and Finance* 25, 1829–1855.
- Werner, I., and A. Kleidon, 1996, UK and US trading of British cross-listed stocks: An intraday analysis of market integration, *Review of Financial Studies* 9, 619–664.
- White, H., and I. Domowitz, 1984, Non-linear regression with dependent observations, *Econometrica* 52, 143–161.
- Wood, R., McNish, T., and J. Ord, 1985, An investigation of intraday data for NYSE stocks, *Journal of Finance* 40, 723–741.
- Yaday, P., and P. Pope, 1990, Stock index futures arbitrage: International evidence, *Journal of Futures Markets* 10, 573–603.
- , 1994, Stock index futures mispricing: Profit opportunities or risk premia? *Journal of Banking and Finance* 18, 921–953.

- Yadav, P., Pope, P., and K. Paudyal, 1994, Threshold autoregressive modeling in finance: The price differences of equivalent assets, *Mathematical Finance* 4, 205–221.
- Zhang, M., Russell, J., and R. Tsay, 2001, A nonlinear autoregressive conditional duration model with applications to financial transaction data, *Journal of Econometrics* 61, 179–207.

Table 1: Empirical test size

This table gives the empirical sizes of various LM tests at various nominal sizes, $\alpha \in \{0.01, 0.05, 0.1\}$, and sample sizes, $T \in \{125, 250, 500, 1000, 10000\}$, using a 10,000 replication Monte Carlo experiment with $IN(0,1)$ errors (Panel A) and bootstrapped errors (Panel B). The test for linearity against the STAR model is given by LM_2 , against the FFF-version of the augmented STAR model is given by LM_{3a} , and against the spline-version of the augmented STAR model is given by LM_{3b} . In addition, the empirical sizes of the tests that smoothness (γ_t) is not dependent on FFF periodic components ($LM_{3a,1}$), and spline periodic components ($LM_{3b,1}$), are given. Finally, small-sample F-distribution versions of these tests are provided (these are denoted by an additional superscript asterisk).

Test	T=125			T=250			T=500			T=1,000			T=10,000		
	$\alpha = 0.01$	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
Panel A: $IN(0,1)$ errors															
LM_2	0.0085	0.0453	0.0915	0.0077	0.0457	0.0955	0.0098	0.0475	0.0985	0.0097	0.0508	0.0987	0.0095	0.0477	0.0962
LM_2^*	0.0086	0.0439	0.0876	0.0077	0.0451	0.0943	0.0098	0.0474	0.0982	0.0097	0.0508	0.0985	0.0095	0.0477	0.0962
LM_{3a}	0.0009	0.0186	0.0558	0.0037	0.0303	0.0742	0.0063	0.0398	0.0824	0.0072	0.0445	0.0913	0.0112	0.0490	0.1044
LM_{3a}^*	0.0055	0.0336	0.0729	0.0064	0.0362	0.0804	0.0072	0.0433	0.0852	0.0083	0.0456	0.0924	0.0114	0.0491	0.1047
$LM_{3a,1}$	0.0016	0.0262	0.0722	0.0051	0.0346	0.0816	0.0061	0.0400	0.0825	0.0088	0.0445	0.0917	0.0121	0.0504	0.1027
$LM_{3a,1}^*$	0.0055	0.0339	0.0729	0.0074	0.0380	0.0811	0.0072	0.0415	0.0821	0.0093	0.0451	0.0916	0.0122	0.0504	0.1027
LM_{3b}	0.0004	0.0118	0.0487	0.0037	0.0299	0.0718	0.0064	0.0378	0.0841	0.0078	0.0445	0.0915	0.0108	0.0502	0.1057
LM_{3b}^*	0.0055	0.0334	0.0710	0.0065	0.0371	0.0807	0.0075	0.0425	0.0871	0.0083	0.0464	0.0926	0.0110	0.0506	0.1061
$LM_{3b,1}$	0.0012	0.0218	0.0682	0.0053	0.0337	0.0828	0.0069	0.0407	0.0862	0.0083	0.0468	0.0938	0.0107	0.0497	0.1019
$LM_{3b,1}^*$	0.0052	0.0343	0.0713	0.0071	0.0363	0.0817	0.0078	0.0421	0.0857	0.0092	0.0472	0.0934	0.0107	0.0498	0.1019
Panel B: Bootstrapped errors															
LM_2	0.0131	0.0495	0.0951	0.0140	0.0521	0.0955	0.0174	0.0523	0.0943	0.0148	0.0530	0.0970	0.0126	0.0498	0.0986
LM_2^*	0.0135	0.0481	0.0915	0.0141	0.0516	0.0932	0.0174	0.0521	0.0938	0.0148	0.0530	0.0964	0.0126	0.0498	0.0986
LM_{3a}	0.0102	0.0427	0.0806	0.0175	0.0550	0.0923	0.0181	0.0565	0.0977	0.0217	0.0598	0.1054	0.0237	0.0644	0.1104
LM_{3a}^*	0.0218	0.0584	0.0970	0.0222	0.0615	0.0976	0.0199	0.0602	0.0998	0.0222	0.0609	0.1070	0.0237	0.0645	0.1110
$LM_{3a,1}$	0.0124	0.0501	0.0993	0.0172	0.0551	0.0986	0.0160	0.0559	0.0956	0.0196	0.0591	0.1050	0.0215	0.0643	0.1098
$LM_{3a,1}^*$	0.0196	0.0570	0.0996	0.0215	0.0575	0.0978	0.0174	0.0573	0.0955	0.0201	0.0593	0.1048	0.0216	0.0644	0.1098
LM_{3b}	0.0055	0.0342	0.0806	0.0162	0.0506	0.0883	0.0186	0.0530	0.0926	0.0211	0.0588	0.1005	0.0222	0.0660	0.1086
LM_{3b}^*	0.0203	0.0629	0.1044	0.0220	0.0572	0.0950	0.0203	0.0562	0.0955	0.0225	0.0602	0.1019	0.0223	0.0661	0.1087
$LM_{3b,1}$	0.0073	0.0473	0.0991	0.0171	0.0516	0.0962	0.0159	0.0511	0.0912	0.0188	0.0572	0.1018	0.0200	0.0615	0.1077
$LM_{3b,1}^*$	0.0195	0.0610	0.1029	0.0203	0.0543	0.0958	0.0172	0.0524	0.0909	0.0195	0.0578	0.1016	0.0200	0.0615	0.1077

Table 2: Testing for linearity

This table gives the results for the LR and LM tests for linearity against the TAR model (LR₁), against the STAR model (LM₂), against the FFF-version of the augmented STAR model (LM_{3a}), and against the spline-version of the augmented STAR model (LM_{3b}). In addition, the results of the test that smoothness (γ_t) is not dependent on FFF periodic components (LM_{3a,1}), and spline periodic components (LM_{3b,1}), are given. These results are based on tests applied to regular futures mispricing (Panel A) and E-mini futures mispricing (Panel B), conducted over the parameter space, $p \in \{1, 2, 3\}$, $d \leq p$, $M = 3$, and $N = 3$. The numbers in parentheses are the p-values associated with the tests.

Test	$p = 1$		$p = 2$		$p = 3$	
	$d = 1$	$d = 1$	$d = 2$	$d = 1$	$d = 2$	$d = 3$
Panel A: Regular futures mispricing						
LR ₁	13.0617 (0.0380)	15.1010 (0.0820)	21.5932 (0.0020)	14.9258 (0.0810)	22.3336 (0.0010)	36.9137 (0.0000)
LM ₂	4.4400 (0.1086)	8.8010 (0.0663)	10.2353 (0.0366)	9.9110 (0.1284)	20.4549 (0.0023)	23.9467 (0.0005)
LM _{3a}	420.8863 (0.0000)	484.0239 (0.0000)	445.2396 (0.0000)	651.4237 (0.0000)	617.5900 (0.0000)	501.1392 (0.0000)
LM _{3a,1}	416.5031 (0.0000)	475.3535 (0.0000)	435.1432 (0.0000)	641.7143 (0.0000)	597.5225 (0.0000)	477.5549 (0.0000)
LM _{3b}	488.9000 (0.0000)	581.1424 (0.0000)	526.4173 (0.0000)	731.7419 (0.0000)	683.5831 (0.0000)	565.6493 (0.0000)
LM _{3b,1}	484.5261 (0.0000)	572.4986 (0.0000)	516.3468 (0.0000)	722.0577 (0.0000)	663.5584 (0.0000)	541.1140 (0.0000)
Panel B: E-mini futures mispricing						
LR ₁	10.8613 (0.1100)	11.4293 (0.1640)	17.8160 (0.0090)	15.7573 (0.0420)	22.6789 (0.0190)	30.3648 (0.0010)
LM ₂	3.6605 (0.1604)	8.8363 (0.0630)	9.7020 (0.0458)	14.3794 (0.0257)	26.4122 (0.0002)	24.5857 (0.0004)
LM _{3a}	421.3527 (0.0000)	511.0843 (0.0000)	460.1240 (0.0000)	681.0049 (0.0000)	623.6115 (0.0000)	505.8192 (0.0000)
LM _{3a,1}	417.7392 (0.0000)	502.2980 (0.0000)	450.5584 (0.0000)	666.9294 (0.0000)	597.6997 (0.0000)	481.6087 (0.0000)
LM _{3b}	487.5700 (0.0000)	619.5107 (0.0000)	545.5842 (0.0000)	776.9611 (0.0000)	694.7807 (0.0000)	575.9093 (0.0000)
LM _{3b,1}	483.9639 (0.0000)	610.7546 (0.0000)	536.0445 (0.0000)	762.9294 (0.0000)	668.9285 (0.0000)	551.7535 (0.0000)

Table 3: Estimation details

This table gives the linear parameter estimates (with associated HCC standard errors given in parentheses), details of the non-linear parameters estimates (with associated HCC standard errors given in parentheses where applicable), and measures of model fit, for the AR model (M0), the TAR model (M1), the STAR model (M2), the FFF-version of the augmented STAR model (M3a), and the spline-version of the augmented STAR model (M3b). These estimates are based on models of regular futures mispricing (Panel A) and E-mini futures mispricing (Panel B), with $p = 2$, $d = 1$, $M = 3$, and $N = 3$. (Note that a confidence interval is given for the size of the arbitrage bound ($\hat{\kappa}$), as measured by MI, as a standard error is not available. Also, the number in parentheses for $\hat{\gamma}_t$ and $\hat{G}(\cdot)$ are the standard deviations of $\hat{\gamma}_t$ and $\hat{G}(\cdot)$, respectively).

Model	Linear Parameters					Non-linear Parameters				Model Fit		
	$\hat{\varphi}_{0,1}$	$\hat{\varphi}_{1,1}$	$\hat{\varphi}_{2,1}$	$\hat{\varphi}_{0,2}^*$	$\hat{\varphi}_{1,2}^*$	$\hat{\varphi}_{2,2}^*$	$\hat{\kappa}$	$\hat{\gamma}$	$\hat{\gamma}_t$	$\hat{G}(\cdot)$	$100 \times \hat{\sigma}_\epsilon$	AIC
Panel A: Regular futures mispricing												
M0	0.0006 (0.0011)	0.9320 (0.0090)	-0.0414 (0.0085)								19.6933	-3.2496
M1	0.0009 (0.0011)	0.9486 (0.0087)	-0.0493 (0.0084)	-0.0025 (0.0051)	-0.0403 (0.0205)	0.0276 (0.0212)	0.6730 [0.44, 0.69]			0.1077 (0.3100)	19.6887	-3.2498
M2	0.0002 (0.0016)	0.9591 (0.0149)	-0.0492 (0.0114)	0.0017 (0.0103)	-0.0452 (0.0283)	0.0185 (0.0284)		2.4707 (0.0957)		0.2471 (0.2708)	19.6892	-3.2497
M3a	0.0018 (0.0015)	0.9860 (0.0135)	-0.0552 (0.0115)	-0.0044 (0.0052)	-0.0960 (0.0238)	0.0339 (0.0230)			5.2464 (4.5848)	0.2973 (0.3380)	19.6589	-3.2523
M3b	0.0018 (0.0015)	0.9858 (0.0133)	-0.0564 (0.0113)	-0.0047 (0.0054)	-0.0985 (0.0241)	0.0374 (0.0232)			4.8511 (4.7365)	0.2817 (0.3306)	19.6593	-3.2522
Panel B: E-mini futures mispricing												
M0	0.0006 (0.0014)	0.9341 (0.0135)	-0.0429 (0.0116)								19.5486	-3.2643
M1	-0.0000 (0.0011)	0.9522 (0.0097)	-0.0452 (0.0091)	0.0025 (0.0029)	-0.0240 (0.0158)	0.0047 (0.0161)	0.4410 [0.00, 4.55]			0.2451 (0.4301)	19.5451	-3.2644
M2	0.0018 (0.0016)	0.9585 (0.0157)	-0.0463 (0.0116)	0.0026 (0.0075)	-0.0337 (0.0282)	0.0067 (0.0282)		3.5607 (0.1265)		0.3027 (0.3016)	19.5455	-3.2643
M3a	0.0018 (0.0015)	0.9826 (0.0127)	-0.0522 (0.0107)	-0.0042 (0.0046)	-0.0842 (0.0239)	0.0232 (0.0227)			6.4329 (6.6613)	0.3037 (0.3486)	19.5165	-3.2668
M3b	0.0023 (0.0014)	0.9848 (0.0126)	-0.0570 (0.0112)	-0.0059 (0.0038)	-0.0920 (0.0214)	0.0337 (0.0204)			11.5493 (15.8934)	0.3238 (0.3760)	19.5137	-3.2670

Table 4: Forecast evaluation

This table gives the test statistics associated with model-based forecast comparisons using the MGN test, the MR test, and the DM test, each based on the AR model (M0), the TAR model (M1), the STAR model (M2), the FFF-version of the augmented STAR model (M3a), and the spline-version of the augmented STAR model (M3b). These comparisons are based on regular futures mispricing (Panel A) and E-mini futures mispricing (Panel B). The numbers in parentheses are the p-values associated with the tests.

Model	MGN Test			MR Test			DM Test					
	M1	M2	M3a	M1	M2	M3a	M1	M2	M3a	M3b		
Panel A: Regular futures mispricing												
M0	1.9570 (0.0504)	1.8342 (0.0666)	5.3757 (0.0000)	5.3468 (0.0000)	1.9461 (0.0516)	1.8246 (0.0681)	5.2349 (0.0000)	5.2667 (0.0000)	1.0818 (0.2793)	1.0695 (0.2848)	2.7765 (0.0055)	2.7439 (0.0061)
M1		-0.3933 (0.6941)	4.9250 (0.0000)	4.9532 (0.0000)		-0.3937 (0.6938)	4.8576 (0.0000)	4.8849 (0.0000)		-0.2797 (0.7797)	2.4397 (0.0147)	2.4414 (0.0146)
M2			5.1843 (0.0000)	5.2083 (0.0000)			5.1094 (0.0000)	5.1325 (0.0000)			2.5317 (0.0113)	2.5301 (0.0114)
M3a				-0.3824 (0.7022)				-0.3829 (0.7018)				-0.2473 (0.8047)
Panel B: E-mini futures mispricing												
M0	1.6826 (0.0925)	1.5898 (0.1119)	5.2041 (0.0000)	5.4315 (0.0000)	1.6745 (0.0940)	1.5825 (0.1135)	5.1287 (0.0000)	5.3495 (0.0000)	1.4268 (0.1536)	1.0479 (0.2947)	2.6716 (0.0075)	2.9574 (0.0031)
M1		-0.3496 (0.7266)	4.9078 (0.0000)	5.0413 (0.0000)		-0.3499 (0.7264)	4.8409 (0.0000)	4.9709 (0.0000)		-0.2667 (0.7897)	2.4740 (0.0134)	2.7108 (0.0067)
M2			5.0797 (0.0000)	5.2011 (0.0000)			5.0080 (0.0000)	5.1261 (0.0000)			2.5379 (0.0112)	2.7370 (0.0062)
M3a				1.3637 (0.1727)				1.3588 (0.1742)				0.5162 (0.6057)

Figure 1: Intraday mispricing and trading volume

This figure shows the intraday mean absolute mispricing associated with regular futures mispricing (Panel A), and E-mini futures mispricing (Panel B), and the respective intraday mean trading volumes (Panel C and Panel D).

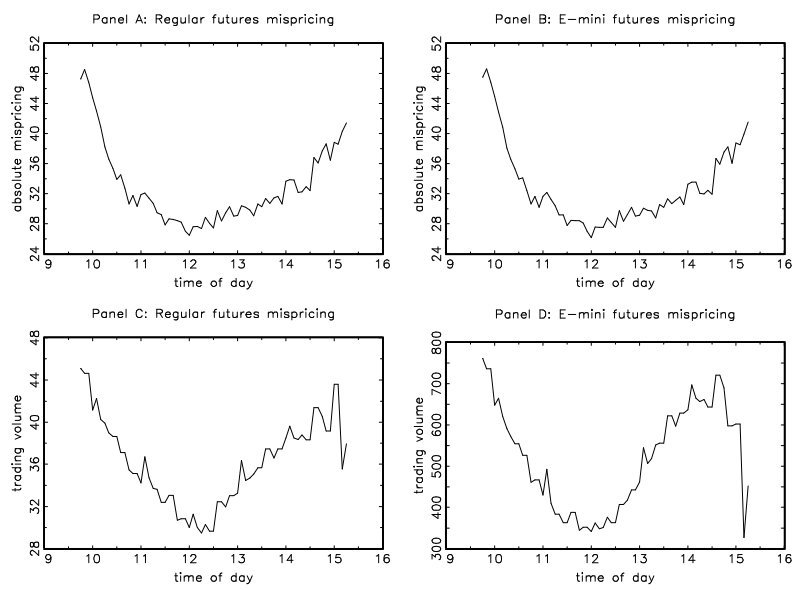


Figure 2: Intraday smoothness and arbitrage activity

This figure shows the intraday propensity-to-arbitrage (smoothness), and the intraday mean level of arbitrage activity (the intraday mean value of the transition function), associated with regular futures mispricing (Panel A and Panel C, respectively), and E-mini futures mispricing (Panel B and Panel D, respectively). These values are calculated using the STAR model (M2), the FFF-version of the augmented STAR model (M3a), and the spline-version of the augmented STAR model (M3b).

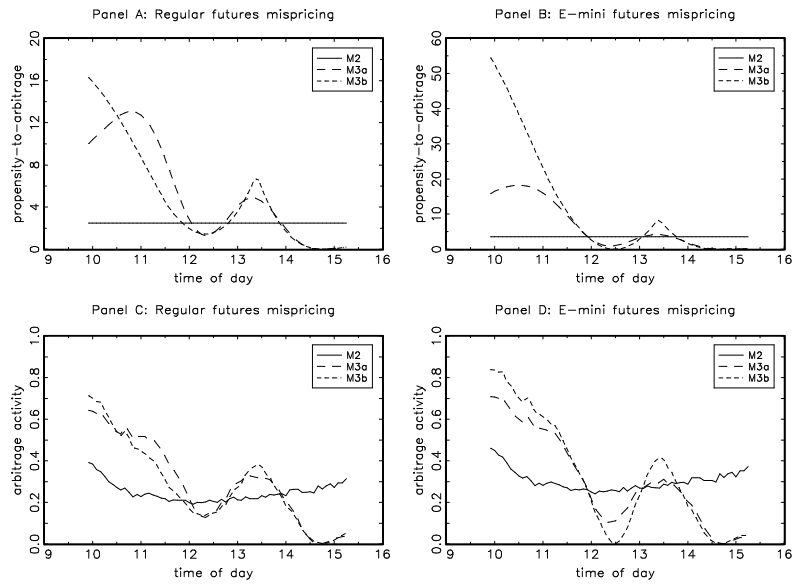


Figure 3: The impact of shocks (regular futures mispricing)

This figure shows the impact of a unit shock to regular futures mispricing. These impacts are measured by the generalized impulse response function, $GIRF_y(h, 1, \omega_{t-1})$, based on the AR model (M0), the STAR model (M2), the FFF-version of the augmented STAR model (M3a), and the spline-version of the augmented STAR model (M3b), for up to 3 hours after the shock occurs ($h \in \{0, 1, 2, \dots, 36\}$). Panel A shows the impact when the shock occurs at any time, Panel B shows the impact when the shock occurs during the first fifteen minutes of trading, and Panel C shows the impact when the shock occurs during the last fifteen minutes of trading. The asymmetry of the shocks, as measured by $ASY_y(h, 1, \omega_{t-1})$, is given in Panel D.

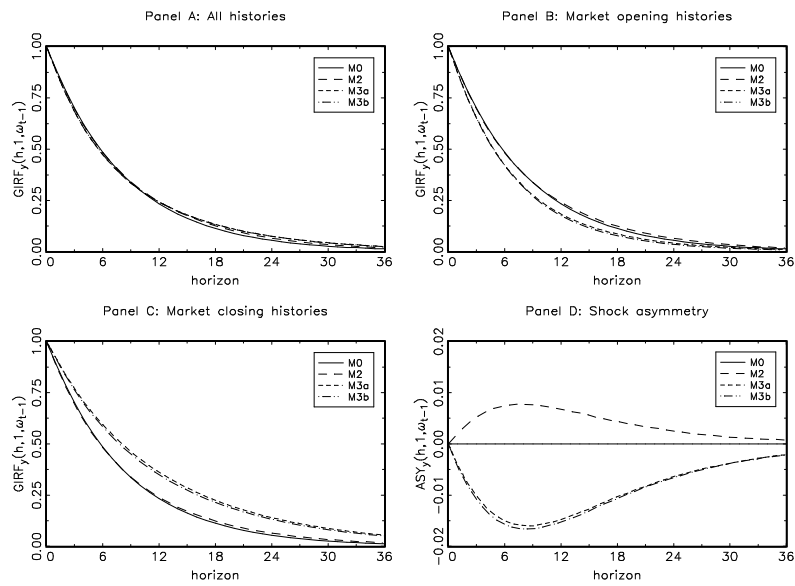


Figure 4: The impact of shocks (E-mini futures mispricing)

This figure shows the impact of a unit shock to E-mini futures mispricing. These impacts are measured by the generalized impulse response function, $GIRF_y(h, 1, \omega_{t-1})$, based on the AR model (M0), the STAR model (M2), the FFF-version of the augmented STAR model (M3a), and the spline-version of the augmented STAR model (M3b), for up to 3 hours after the shock occurs ($h \in \{0, 1, 2, \dots, 36\}$). Panel A shows the impact when the shock occurs at any time, Panel B shows the impact when the shock occurs during the first fifteen minutes of trading, and Panel C shows the impact when the shock occurs during the last fifteen minutes of trading. The asymmetry of the shocks, as measured by $ASY_y(h, 1, \omega_{t-1})$, is given in Panel D.

