



# Testing futures returns predictability: implications for hedgers

Cédric de Ville de Goyet, Geert Dhaene and Piet Sercu

DEPARTMENT OF ACCOUNTANCY, FINANCE AND INSURANCE (AFI)

# Testing futures returns predictability: implications for hedgers

Cédric de Ville de Goyet\*, Geert Dhaene<sup>†</sup> and Piet Sercu<sup>‡</sup>

February 5, 2007

\*F.W.O. (Science Foundation) researcher at KU Leuven, Faculty of Economics and Business Studies, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 326 658; email: cedric.deville@econ.kuleuven.be.

<sup>†</sup>KU Leuven, Faculty of Economics and Business Studies, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 326798; email: geert.dhaene@econ.kuleuven.be.

<sup>‡</sup>KU Leuven, Faculty of Economics and Business Studies, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 326756; email: piet.sercu@econ.kuleuven.be.

## **Abstract**

The predictability of futures returns is investigated using a semiparametric approach where it is assumed that the expected returns depend nonparametrically on a combination of predictors. We first collapse the forecasting variables into a single-index variable where the weights are identified up to scale, using the average derivative estimator proposed by Stoker (1986). We then use the Nadaraya-Watson kernel estimator to calculate (and visually depict) the relation between the estimated index and the expected futures returns. An application to four agricultural commodity futures illustrates the technique. The results indicate that for each of the commodities considered, the estimated index contains statistically significant information regarding the expected futures returns. Economic implications for a non-infinitely risk averse hedger are also discussed.

JEL classification: G11, G14, C14.

Key words: hedging, futures market, average derivative estimates.

# 1 Introduction

Under the assumptions that agents are risk neutral, have common and constant preferences, and are rational, Samuelson (1965) proves that futures prices fluctuate randomly.<sup>1</sup> To see the consequence for the hedger, consider an agent with a long position in the spot market who wishes to place a hedge using a futures contract. The relative size of the short position in the futures market, also called the hedge ratio, has to be determined in some optimal way. The conventional approach followed in much of the hedging literature is to fully hedge by solely focusing on the risk associated with the random portfolio return. The optimal hedge ratio is then chosen to minimize the variance of the portfolio return (see Johnson, 1960; Stein, 1961; Ederington, 1979; or Baillie and Myers, 1991). Some authors, however, have incorporated both risk and return in their hedging policy (see for example Hsin, Kuo and Lee, 1994). The optimal hedge ratio consistent with the mean-variance framework then is the sum of two components. The first component, a pure hedging term, is the conventional minimum-variance hedge ratio and the second, the speculative part, is a function of the expected futures return, the risk aversion parameter and the conditional variance of the futures return. If the hedger is infinitely risk averse or if the futures prices fluctuate randomly, the speculative component disappears and the mean-variance hedge ratio reduces to the minimum-variance hedge ratio. When the agent's degree of risk aversion is not too high and when the futures prices are predictable, the agent may wish to exploit the bias in an attempt to trade-off risk against return. In other words, rejection of Samuelson's hypothesis indicates a possible risk/return tradeoff in hedging.

The purpose of this paper is to test empirically whether commodity futures prices fluctuate randomly and discuss the implication of predictability for hedgers. Bessembinder and Chan (1992) and Miffre (2002) report predictability in various futures markets. They use linear models with macroeconomic predictors such as the term structure of interest rates, inflation, the difference between low- and high-grade bond yields etc. McCurdy and Morgan (1988) reject the martingale hypothesis for the Deutsche Mark futures. The approach adopted here differs from previous research as it relies on a nonlinear model and basically uses the lagged underlying spot return, basis and futures return as predictors. There is also more focus on economic relevance besides statistical significance, in the sense that we measure the impact on the hedge ratio.

---

<sup>1</sup>See also LeRoy (1989) or Campbell, Lo and MacKinley (1996, p 23-24) for a lucid review of the proof.

To address the predictability issue, we use a semiparametric approach where the expected return depends nonparametrically on a parametric index. More precisely, we optimally combine several variables to predict the next period futures return. This enables us to judge the importance of each individual variable in the index with minimal a priori restrictions on the parametric functional form: it only requires expected futures returns to be weakly dependent and to be functionally related, in a time invariant way, to a linear index of the conditioning variables. The relation between the expected return and the index is left unspecified and may well be highly nonlinear.

In terms of forecasting methodology, two steps are taken. We first collapse the forecasting variables into a single-index variable where the optimal weights are identified up to scale, using the average derivative estimator (ADE) proposed by Stoker (1986). Robust Wald and  $t$ -tests are also performed on the estimated weights in order to analyze the impact of the predictors on the expected futures returns. We then use the Nadaraya-Watson kernel estimator to calculate and visually depict the relation between the estimated index and the expected futures returns.

For corn, wheat, soybeans and oats, our results show that the constructed index contains information for predicting futures returns on a daily horizon. The paper closes by assessing the relevance of this predictability for the hedger. Our conclusions indicate that, from the hedger's point of view, there is a risk/return tradeoff in hedging and that the mean-variance optimal hedge ratio is substantially affected, even though there is a high uncertainty surrounding it.

The remainder of this paper is organized as follows. Section 2 deals with the methodology. We motivate the use of the nonparametric Nadaraya-Watson kernel estimator to test the predictability of futures returns and discuss its applicability when returns are related to several predictors. The ADE is then used to collapse the many variables into a single index. We end the section by indicating how to measure the economic relevance of the prediction for hedgers. Section 3 applies the proposed methodology to real agricultural data. We first test predictability of the futures returns and then discuss the practical implications for the hedger. Section 4 concludes.

## 2 Methodology

### 2.1 A nonparametric analysis of return predictability

Let  $s_t$  and  $f_t$  denote the log price of the spot and the futures contract nearest to maturity, respectively. We test the hypothesis that futures returns are conditionally unpredictable:

$$H_{0,\mathbf{x}} : E[\Delta f_t | \mathbf{x}_{t-1}] = 0,$$

where  $\mathbf{x}_{t-1} \equiv (x_{1,t-1}, \dots, x_{q,t-1})'$  is a  $q \times 1$  vector of explanatory variables, available to the agent at time  $t - 1$ . We consider a nonparametric regression method to estimate  $E[\Delta f_t | \mathbf{x}_{t-1}] \equiv m(\mathbf{x}_{t-1})$ , where  $m(\cdot)$  is an unknown, possibly nonlinear, scalar-valued function, and we test the hypothesis that  $m(\cdot)$  is identically zero.

The idea behind the nonparametric technique is to replace the conditional expectation with a local sample average of the futures returns. When  $\mathbf{x}$  varies continuously, a general class of nonparametric estimators of the conditional mean  $E[\Delta f_t | \mathbf{x}_{t-1} = \mathbf{x}]$  can be written as

$$\hat{m}(\mathbf{x}) \equiv T^{-1} \sum_{t=1}^T \omega\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h}\right) \Delta f_t, \quad (2.1)$$

where  $\omega(\cdot)$  is a weight function, assigning weights to the observations, and  $h$  is a bandwidth or smoothing parameter. The weight function is chosen such that it gives more importance to observations where  $\mathbf{x}_{t-1}$  is close to  $\mathbf{x}$ . Following standard practice, we adopt the Nadaraya-Watson kernel estimator by substituting

$$\omega\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h}\right) \equiv \frac{K\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h}\right)}{T^{-1} \sum_{t=1}^T K\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h}\right)}$$

into (2.1), where  $K(\cdot)$  is a  $q$ -variate kernel. In the application, we use the Gaussian kernel<sup>2</sup>

$$K(u_1, \dots, u_q) = \prod_{i=1}^q \frac{1}{\sqrt{2\pi}} e^{-\frac{u_i^2}{2}}.$$

It is well known that, under general assumptions,  $\hat{m}(\mathbf{x})$  is a consistent estimate of the true conditional expectation,  $m(\mathbf{x}) = E[\Delta f_t | \mathbf{x}_{t-1} = \mathbf{x}]$ .

The choice of the bandwidth  $h$  is of great importance. It regulates the size of the neighborhood around  $\mathbf{x}$ : if  $h$  is very large, the estimate  $\hat{m}(\mathbf{x})$  is computed over a large neighborhood,

---

<sup>2</sup>The choice of a “best” kernel is less important than the choice of the “best” bandwidth (see, e.g., Pagan and Ullah, 1999). What makes the Gaussian kernel attractive is its continuity and the amount of differentiability it possesses.

thus giving importance to observations where  $\mathbf{x}_{t-1}$  is far away from  $\mathbf{x}$ , inducing bias in the estimator. In contrast, if  $h$  is very small,  $\hat{m}(\mathbf{x})$  is computed over a small neighborhood, which increases the variance of the estimator. The optimal choice of  $h$  should balance bias against variance. We use the leave-one-out cross-validation method to obtain  $h$ .<sup>3</sup> Essentially, this is a data-driven bandwidth selection procedure that minimizes the mean squared error of the estimates computed in an out-of-sample setting (see Härdle (1990), Chapter 5).

Testing the null  $H_{0,\mathbf{x}}$  can be easily done by computing the 95% pointwise confidence bounds around  $\hat{m}(\mathbf{x})$ , for all  $\mathbf{x}$ . The bounds are derived as follows: the stationary block-bootstrap of Politis and Romano (1994) is applied to the vector of futures returns and predictors,  $(\Delta f_t, \mathbf{x}_{t-1})$ , jointly. Essentially, the block-bootstrap builds series from resampled blocks (of random length) of the original data.<sup>4</sup> This preserves serial correlation, GARCH effects and non-normalities in the data. For each bootstrapped bivariate series, the Nadaraya-Watson estimator  $\hat{m}(\mathbf{x})$  is computed, giving the bootstrap distribution of  $\hat{m}(\mathbf{x})$ . The bounds follow as the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the bootstrap distribution.

In practice, we may wish to relate  $\Delta f_t$  to many predictors. As predictors we consider

$$\mathbf{x}_{t-1} = \left( \Delta f_{t-1}, \Delta s_{t-1}, (\Delta f_{t-1})^2, (\Delta s_{t-1})^2, \Delta f_{t-1} \Delta s_{t-1}, f_{t-1} - s_{t-1} \right)'$$

However, the nonparametric method applied to estimate conditional expectations may not be able to handle a large number of conditioning variables, a problem known as “the curse of dimensionality”. One technique for dimension reduction is to impose restrictions on the types of nonlinearities one is dealing with. One may assume, for instance, that a linear combination of the conditioning variables, a so-called index variable  $\mathbf{z}_{t-1} \equiv \mathbf{x}'_{t-1} \boldsymbol{\beta}$ , is nonparametrically related to the futures returns. In this case, we get rid of the curse of dimensionality problem (as the nonparametric estimator becomes a function of an univariate predictor, the linear index itself) and still allow for important nonlinearities in the relations between  $\mathbf{x}'_{t-1} \boldsymbol{\beta}$  and  $\Delta f_t$ . This enables us to track the predictors with the highest information content and to combine them in an optimal way. The major issue then becomes the estimation of  $\boldsymbol{\beta}$  since, on that assumption, we can test the null  $H_{0,\mathbf{x}}$  conditional on the one-dimensional estimated index variable  $\hat{\mathbf{z}}_{t-1}$  by substituting  $\mathbf{x}_{t-1}$  by  $\hat{\mathbf{z}}_{t-1}$  in (2.1). The estimation of  $\boldsymbol{\beta}$  (up to scale) is the object of the

---

<sup>3</sup>Note that we select a unique bandwidth  $h$  for all  $q$  predictors. This is reasonable given that we standardize all the predictors.

<sup>4</sup>The length of each block is drawn independently from a geometric distribution. Politis and Romano (1994) further specify that the original data have to be “wrapped” to ensure that whenever a block goes past the last observation, it can be filled with observations from the beginning of the original series.

following section.

## 2.2 Semiparametric estimation of the index

In this section, we first discuss the estimation of the vector  $\beta$  in an i.i.d. context using the average derivative technique. We end the section by indicating how it can be applied to times series.

### 2.2.1 Average derivative estimators

Consider a model where the conditional expectation of futures returns may be written in the single-index form

$$E[\Delta f_t | \mathbf{x}_{t-1}] = m(\mathbf{z}_{t-1}). \quad (2.2)$$

Without specific assumptions on  $m$ , Ichimura (1993) searches over a large grid the  $\hat{\beta}$  that minimizes the sum of the squares of the residuals  $\Delta f_t - \hat{m}(\mathbf{x}'_{t-1}\beta)$ . This technique enjoys very good asymptotic properties but is hard to implement because the objective function may not be concave or unimodal. Alternatively, Stoker (1986) notes that the  $q \times 1$  vector of coefficients  $\beta$  can be estimated up to scale via the derivatives of  $m$  with respect to  $\mathbf{x}_{t-1}$ . This works as follows.

Let  $\delta$  be the vector of average derivatives of  $\Delta f_t$  on  $\mathbf{x}_{t-1}$ , *i.e.* the mean of the  $q \times 1$  vector of partial derivatives  $\frac{\partial m}{\partial \mathbf{x}_{t-1}}$  over the distribution of  $\mathbf{x}_{t-1}$ . By exploiting the chain-rule of differentiation, Stoker (1986) observes that

$$\delta \equiv E\left[\frac{\partial m}{\partial \mathbf{x}_{t-1}}\right] = E\left[\frac{dm}{d\mathbf{z}_{t-1}}\right] \beta = \gamma \beta, \quad (2.3)$$

where  $\gamma \equiv E[dm/d\mathbf{z}_{t-1}]$  is a scalar, assumed different from zero. Equation (2.3) says that  $\delta$ , the vector of average derivatives, is proportional to  $\beta$ , the vector of the coefficients. The scaling coefficient  $\gamma$  is not important here as it can be absorbed into  $m(\cdot)$ . Hence, a consistent estimator of  $\delta$  measures  $\beta$  up to a scale factor.

Let

$$\hat{g}(\mathbf{x}) = T^{-1} h_{RP}^{-q} \sum_{t=1}^T K\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h_{RP}}\right)$$

and

$$\hat{g}'(\mathbf{x}) = T^{-1} h_{RP}^{-q-1} \sum_{t=1}^T K'\left(\frac{\mathbf{x} - \mathbf{x}_{t-1}}{h_{RP}}\right)$$



be the Rosenblatt-Parzen estimators<sup>5</sup> of the marginal density of  $\mathbf{x}$  and its derivative, respectively. We focus on the estimated negative of the score vector,  $\hat{s}(\mathbf{x}) \equiv -\hat{g}'(\mathbf{x})/\hat{g}(\mathbf{x}) = -\partial \ln \hat{g}(\mathbf{x})/\partial \mathbf{x}$ .

Several consistent ADEs are available in the literature. A simple direct estimator of  $\delta$  is given by the sample analog to (2.3):

$$\hat{\delta}^{dir} = T^{-1} \sum_{t=1}^T \hat{m}'(\mathbf{x}_{t-1}), \quad (2.4)$$

where  $\hat{m}'(\mathbf{x}_{t-1})$  is an estimate of the  $q$ -vector of partial derivatives  $\frac{\partial m}{\partial \mathbf{x}_{t-1}}$ .

Two alternative estimators of interest follow from the formula in Stoker (1986, Theorem 1) which connects the average derivative  $\delta$  to the scores  $s(\mathbf{x})$ :

$$\delta = \text{cov}(s(\mathbf{x}), \Delta f), \quad (2.5)$$

$$= \{\text{cov}(s(\mathbf{x}), \mathbf{x})\}^{-1} \text{cov}(s(\mathbf{x}), \Delta f), \quad (2.6)$$

where the second equality derives from the fact that  $\text{cov}(s(\mathbf{x}), \mathbf{x})$  equals a  $q \times q$  identity matrix.<sup>6</sup> Each of these two equations provides an estimator for  $\delta$ . The first equality (2.5) motivates the indirect one. Following Härdle and Stoker (1989),

$$\hat{\delta}^{ind} = T^{-1} \sum_{t=1}^T \hat{s}(\mathbf{x}_{t-1}) \Delta f_t 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]}, \quad (2.7)$$

where  $1_{[\cdot]}$  is the indicator function and  $b$  is chosen so that some chosen fraction (between 1% and 5%) of the observations are dropped. This “trimming” procedure drops observations with a very small estimated density  $\hat{g}(\mathbf{x})$ , the reason being that  $\hat{s}(\mathbf{x})$  may not behave well when  $\hat{g}'(\mathbf{x})$  is divided by a very small  $\hat{g}(\mathbf{x})$ . The second equality (2.6) points out that the ADE may be viewed as the instrumental variable slope coefficient of a standard linear regression of  $\Delta f_t$  on  $\mathbf{x}_{t-1}$  with the trimmed scores as the instrumental variable, suggesting the estimator

$$\hat{\delta}^{IV} = \left( T^{-1} \sum_{t=1}^T \hat{s}(\mathbf{x}_{t-1}) \mathbf{x}'_{t-1} 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]} \right)^{-1} \left( T^{-1} \sum_{t=1}^T \hat{s}(\mathbf{x}_{t-1}) \Delta f_t 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]} \right). \quad (2.8)$$

---

<sup>5</sup>The bandwidth of the Rosenblatt-Parzen estimator  $h_{RP}$  may be different from the bandwidth of the Nadaraya-Watson estimator  $h$ . As the kernel is multivariate standard normal and as we standardized the predictors beforehand in the application, we use the popular plug-in estimate of Silverman (1986):  $h_{RP} = 1.06T^{-1/5}$ . The same bandwidth is used for both estimators  $\hat{g}(\mathbf{x})$  and  $\hat{g}'(\mathbf{x})$ .

<sup>6</sup>Assuming that  $g(x)$  vanishes when  $x \rightarrow \pm\infty$  and that  $m(x)$  is bounded, we have:  $E\left[\frac{\partial m(x)}{\partial x}\right] = \int_{-\infty}^{+\infty} \frac{\partial m(x)}{\partial x} g(x) dx = [m(x)g(x)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{\partial g(x)}{\partial x} m(x) dx = -\int_{-\infty}^{+\infty} \frac{\partial g(x)}{\partial x} m(x) g(x) g^{-1}(x) dx = E[m(x)s(x)]$ . Equation (2.5) follows by noting that  $E[s(x)]$  is zero. Similarly, equation (2.6) follows from  $\text{cov}(s(x), x) = E[s(x)x] = -\int_{-\infty}^{+\infty} \frac{\partial g(x)}{\partial x} g^{-1}(x) x g(x) dx = \int_{-\infty}^{+\infty} g(x) dx$ .

Stoker (1993) shows that the three estimators (2.4), (2.7) and (2.8) are  $\sqrt{T}$ -consistent<sup>7</sup> and asymptotically normal. Moreover, they are first-order equivalent<sup>8</sup> to each other (Li, 1996). Stoker (1993) favors  $\hat{\boldsymbol{\delta}}^{IV}$  since it allows its nonparametric components to be less accurately estimated,<sup>9</sup> as the ratio form of the estimator partially cancels out inaccuracy.

Under some general conditions, Härdle and Stoker (1989, Theorem 3.1, Equation (3.6)) show that a consistent estimator of  $\boldsymbol{\Sigma} \equiv \lim_{T \rightarrow \infty} \text{var} \left[ \sqrt{T} \left( \hat{\boldsymbol{\delta}}^{ind} - \boldsymbol{\delta} \right) \right]$  is given by

$$\hat{\boldsymbol{\Sigma}} = T^{-1} \sum_{t=1}^T \hat{\mathbf{r}}_t \hat{\mathbf{r}}_t' 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]} - \bar{\mathbf{r}}_t \bar{\mathbf{r}}_t',$$

where

$$\begin{aligned} \hat{\mathbf{r}}_t = & \hat{s}(\mathbf{x}_{t-1}) \Delta f_t 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]} + T^{-1} h_{RP}^{-k} \sum_{t=1}^T \left[ h_{RP}^{-1} K' \left( \frac{\mathbf{x} - \mathbf{x}_{t-1}}{h_{RP}} \right) \right. \\ & \left. - K \left( \frac{\mathbf{x} - \mathbf{x}_{t-1}}{h_{RP}} \right) \hat{s}(\mathbf{x}_{t-1}) \right] \frac{\Delta f_t 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]}}{\hat{g}(\mathbf{x}_{t-1})} \end{aligned} \quad (2.9)$$

and  $\bar{\mathbf{r}}_t = T^{-1} \sum_{t=1}^T \hat{\mathbf{r}}_t 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]}$ . Hence, the covariance matrix of  $\hat{\boldsymbol{\delta}}^{ind}$  is estimated by  $T^{-1} \hat{\boldsymbol{\Sigma}}$ . Recalling that  $\text{cov}(s(\mathbf{x}), \mathbf{x}) = 1$ , it is obvious that

$$\sqrt{T} \left( \hat{\boldsymbol{\delta}}^{IV} - \boldsymbol{\delta} \right) \xrightarrow{d} N(0, \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}'),$$

where  $\mathbf{A} \equiv \{\text{cov}(s(\mathbf{x}), \mathbf{x})\}^{-1}$ . Thus, the covariance matrix of  $\hat{\boldsymbol{\delta}}^{IV}$  may be estimated by  $T^{-1} \hat{\mathbf{A}} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{A}}$  with  $\hat{\mathbf{A}} \equiv \left\{ T^{-1} \sum_{t=1}^T \hat{s}(\mathbf{x}_{t-1}) \mathbf{x}'_{t-1} 1_{[\hat{g}(\mathbf{x}_{t-1}) > b]} \right\}^{-1}$ .

### 2.2.2 Time series issues

The theory for average derivative estimation applies in situations where the regression errors  $\Delta f_t - \mathbf{x}'_{t-1} \boldsymbol{\delta}$  and the observations are *i.i.d.* However, the kernel density estimators which are the ingredients needed to build the ADEs are still consistent and asymptotically normal under the weaker assumption that the data are stationary and  $\alpha$ -mixing<sup>10</sup> (Pagan and Ullah (1999),

---

<sup>7</sup>Remark that we get a (parametric)  $\sqrt{T}$  rate of convergence for the ADEs while the ingredients building this estimator,  $\hat{g}(\cdot)$  and  $\hat{g}'(\cdot)$ , achieve consistency at a slower rate. In fact, averaging reduces the variance of the estimator. See Ichimura and Todd (2006, p. 77).

<sup>8</sup>Two estimator,  $\hat{\delta}_1$  and  $\hat{\delta}_2$ , are first-order equivalent if  $\sqrt{T} \left( \hat{\delta}_1 - \hat{\delta}_2 \right) \xrightarrow{p} 0$ .

<sup>9</sup>There is a systematic downward bias in density kernel derivative estimators  $\hat{g}'(\mathbf{x})$ . The same bias applies to the estimated scores, as  $\hat{s}(\mathbf{x}) \equiv -\hat{g}'(\mathbf{x}) / \hat{g}(\mathbf{x})$ ; see Stoker (1993).

<sup>10</sup>An  $\alpha$ -mixing series satisfies a condition on the speed at which the influence of the past events on future events disappears, as the interval of time between past and future events increases.

Chapter 2). If so, the ADEs remain consistent<sup>11</sup> and asymptotically normal with weakly dependent observations.

The covariance matrix  $T^{-1}\hat{\mathbf{A}}\hat{\Sigma}\hat{\mathbf{A}}$  has to be corrected to account for heteroskedasticity and autocorrelation in  $\Delta f_t$ . A heteroskedasticity and autocorrelation consistent covariance matrix is given by  $T^{-1}\hat{\mathbf{A}}\hat{\Sigma}^{NW}\hat{\mathbf{A}}$  where  $\hat{\Sigma}^{NW}$  is the Newey-West estimator of the long-run covariance matrix of  $\hat{\mathbf{r}}_t$ .<sup>12</sup>

### 2.3 Measuring economic relevance

While we can easily establish whether and when a non-zero forecast is statistically significant, there is no unique measure of economic relevance. The criterion we adopt is the size of the impact on the hedge ratios. The risk-averse hedger maximizes a mean-variance objective function of the return.<sup>13</sup> The time-varying optimal mean-variance hedge ratio that incorporates both risk and return is given by

$$\begin{aligned} h_{t-1}^{\mu\sigma^2} &\equiv \arg \max_{h_{t-1}} \left[ E[\Delta p_t | \mathbf{x}_{t-1}] - \frac{A}{2} \text{var}[\Delta p_t | \mathbf{x}_{t-1}] \right], \\ &= \frac{\text{cov}(\Delta s_t, \Delta f_t | \mathbf{x}_{t-1}) - \frac{1}{A} E[\Delta f_t | \mathbf{x}_{t-1}]}{\text{var}(\Delta f_t | \mathbf{x}_{t-1})}, \end{aligned} \quad (2.10)$$

where  $\Delta p_t \equiv \Delta s_t - h_{t-1} \Delta f_t$  is the portfolio return and  $A$  is the relative risk aversion parameter. This mean-variance hedge ratio  $h_{t-1}^{\mu\sigma^2}$  in (2.10) reduces to the minimum-variance hedge ratio,

$$\begin{aligned} h_{t-1}^{MinVar} &\equiv \arg \min_{h_{t-1}} \text{var}[\Delta p_t | \mathbf{x}_{t-1}], \\ &= \frac{\text{cov}(\Delta s_t, \Delta f_t | \mathbf{x}_{t-1})}{\text{var}(\Delta f_t | \mathbf{x}_{t-1})}, \end{aligned} \quad (2.11)$$

if the hedger is infinitely risk averse, *i.e.*  $A \rightarrow \infty$ , or if the expected return on the futures contract  $E[\Delta f_t | \mathbf{x}_{t-1}]$  is zero. So, when considering non-infinitely risk-averse hedgers, pre-

---

<sup>11</sup>Ghysels and Ng (1998) claim that  $\hat{\delta}^{IV}$  remains  $\sqrt{T}$ -consistent and asymptotically normal invoking the paper of Chen and Shen (1998) where the  $\sqrt{T}$ -convergence rate and asymptotic normality of the Sieve extremum class of estimates (SEE) is obtained for weakly dependent data. In fact, the claim seems incomplete as  $\hat{\delta}^{IV}$  cannot be written as a SEE. At the very most, it can be seen as an extremum estimator (in the sense defined in Hayashi (2000), Chapter 7) as  $\hat{\delta}^{IV}$  is a variant of a GMM estimate with  $\hat{s}(\mathbf{x})$  as the instrument. However, such an extremum estimator does not belong to the class of the SEE.

<sup>12</sup>We set the lag truncation parameter at  $\left\lfloor 4(T/100)^{2/9} \right\rfloor$ , a common choice.

<sup>13</sup>See Lien and Tse (2002) or Chen, Lee and Shrestha (2003) for a general review on the hedging literature. The obvious shortcoming of this model is that, formally, it assumes that there is just one source of risk to be hedged, and one hedge. However, in view of the low correlations between returns from commodities and financial assets, a true multi-risk solution would come up with essentially the same hedge strategy.

dictability or no predictability of the futures prices bears on choice between the mean- and minimum-variance hedge ratio.

The mean-variance model takes a portfolio point of view, where the producer takes the payoffs of the assets as given and merely tries to combine them optimally. In reality, corporate hedging also affects the payoffs if it reduces the expected costs of financial distress and so on. Since we cannot possibly include this consideration into the optimal hedge, we cannot estimate the optimal hedge. Still, we can compute by how much the optimal hedge is changed by any predictability for a given operational effect: its differential effect on the hedge ratio is measured by

$$\Delta h_{t-1} \equiv h_{t-1}^{\mu\sigma^2} - h_{t-1}^{MinVar} = \frac{-\frac{1}{A} E[\Delta f_t | \mathbf{x}_{t-1}]}{\text{var}(\Delta f_t | \mathbf{x}_{t-1})}. \quad (2.12)$$

Thus, a sensible measure of relevance is the ratio of excess expected return<sup>14</sup> to variance, not standard deviation. The standard deviation would have been adopted in a Sharpe ratio, a criterion that is popular among traders even if theoretically it only makes sense for entire portfolios, not parts thereof. We consider the impact of predictability on the hedge ratio for measures of relative risk aversion ranging between two and ten.

### 3 Application to agricultural futures

We now turn to an empirical application of the technique suggested. We start by describing the data and discuss the potential spurious predictability arising from the fact that spot and futures prices are not synchronized. Then we estimate the regression coefficients, graph each estimated conditional expectation given the constructed index and end the section by discussing the empirical implications for hedgers.

#### 3.1 Data

The data consist of daily spot and nearest-to-maturity futures closing prices of corn, wheat, soybeans and oats. Daytime futures closing prices were extracted from the Chicago Board of Trade tapes and cover the period 1 January 1979 through 31 December 2003 (23 years), for all four commodities considered. For the same period, the commodity spot prices were extracted from Datastream. Corn, soybeans and wheat spot prices (in cents/bushel) are the average of

---

<sup>14</sup>Since a futures contract is a zero-investment instrument, its percentage change is similar to an excess return for a regular asset.

all Central Illinois Elevators prices paid by the Country Elevators to the producers, after 2:00 p.m. Oats spot prices are from Minneapolis, Minnesota. The spot prices are for the following qualities: oats, No.2; wheat, No.2, Soft Red; soybeans, No.1, Yellow; corn, No.2, Yellow.

Even though both states, Illinois and Minnesota, are in the same time zone, futures returns and the lagged spot return predictors are overlapping. The reason is that the agricultural futures market closes before 1:15 pm for the period under consideration, so that their closing prices miss possible information available after 2:00 pm, the afternoon when spot prices are set. Suppose we want to predict the Wednesday-Thursday change in the futures price, noon to noon. Ideally, we should use tuesday-wednesday changes in the spot price, noon to noon. In reality we only have an afternoon-to-afternoon spot return for Tuesday-Wednesday, which includes information not yet available when the prediction for the futures market is to be made and which therefore overestimates the predictive power of the spot return. On the other hand, the alternative of using an afternoon-to-afternoon spot return for Monday-Tuesday would underestimate the predictability since all information that became available Tuesday evening and Wednesday morning is ignored. For our purpose, erring in the conservative side is preferable to the alternative, so our predictions are based on just price information that is available. That is, we use the following predictors:

$$\mathbf{x}_{t-1}^* = \left( \Delta f_{t-1}, \Delta s_{t-2}, f_{t-2} - s_{t-2}, (\Delta f_{t-1})^2, (\Delta s_{t-2})^2, \Delta f_{t-2} \Delta s_{t-2} \right)'$$

For completeness, however, we also re-estimate everything with  $\mathbf{x}_{t-1}$ , ignoring the synchronization issue. This at least provides an indirect indication as to whether the information ignored by  $\mathbf{x}_{t-1}^*$  is very important and whether the predictability pattern is seriously affected.

Following standard practice in the literature (see for example Bera, Garcia and Roh (1997), Harris and Shen (2003)), and in order to avoid thin trading and expiration effects, a contract that expires in month  $m$  is replaced with the next expiring contract on the last day of month  $m - 1$ . Specifically, on the last day of  $m - 1$ ,  $\Delta f_t$  is set equal to the return on the former contract, while on the first day of  $m$ ,  $\Delta f_t$  is set equal to the return on the latter contract.

The prediction horizons considered in this paper are one day and one week. Similar horizons have been considered in the literature; see for instance Baillie and Myers (1991), Bera, Garcia and Roh (1997), Lien, Tse and Tsui (2002) or Byström (2003). The returns were aggregated to yield weekly (Friday to Friday) returns. Table 1 gives descriptive statistics on  $\Delta s_t$  and  $\Delta f_t$  for each commodity on the 23-year data span. It shows the mean, standard deviation, skewness, kurtosis and autocorrelations. The returns data are non-normal as evidenced by the

Table 1: Descriptive statistics on spot and futures returns

	mean	std. dev.	skew.	kurt.	autocorrel. coef.		
	$\times 10^2$				$\rho_1$	$\rho_6$	$\rho_{12}$
Corn, daily, 2/1/1979-31/12/1993							
$\Delta s_t$	0.006	0.0144	-0.483	8.977	0.019	0.022	-0.004
$\Delta f_t$	-0.290	0.0122	-0.006	5.656	0.050	-0.007	0.010
Corn, weekly, 19/1/1979-31/12/1993							
$\Delta s_t$	-0.015	0.0325	-0.496	7.219	0.040	0.027	-0.037
$\Delta f_t$	-1.433	0.0280	0.301	6.826	-0.016	-0.011	-0.000
Wheat, daily, 4/1/1983-31/12/1993							
$\Delta s_t$	-0.007	0.0172	-0.973	22.108	-0.008	0.004	0.022
$\Delta f_t$	-0.205	0.0137	-0.814	5.539	0.024	0.003	0.021
Wheat, weekly, 6/1/1984-31/12/1993							
$\Delta s_t$	0.000	0.0372	-0.661	10.973	-0.020	-0.018	0.018
$\Delta f_t$	-1.003	0.0301	0.423	4.679	0.002	-0.031	0.026
Oats, daily, 3/1/1979-31/12/1993							
$\Delta s_t$	0.027	0.0193	-0.079	22.967	-0.029	0.011	-0.001
$\Delta f_t$	0.341	0.0178	-0.057	5.127	0.057	-0.010	-0.004
Oats, weekly, 19/1/1979-31/12/1993							
$\Delta s_t$	0.065	0.0420	-0.056	7.703	-0.071	-0.032	-0.026
$\Delta f_t$	-1.727	0.0415	0.131	7.017	-0.051	-0.019	0.041
Soybeans, daily, 3/1/1979-31/12/1993							
$\Delta s_t$	-0.023	0.0135	-0.372	6.594	-0.029	-0.021	-0.011
$\Delta f_t$	-0.162	0.0129	-0.151	5.306	-0.017	-0.023	0.003
Soybeans, weekly, 12/1/1979-31/12/1993							
$\Delta s_t$	0.084	0.0307	-0.144	6.089	-0.087	-0.012	-0.013
$\Delta f_t$	0.805	0.0294	-0.159	6.311	-0.054	-0.002	-0.031

high kurtosis and the significant Jarque-Bera statistics (not reported here).

### 3.2 Predicting futures returns

We first predict  $\Delta f_t$  using  $\mathbf{x}_{t-1}^*$ . Panels A and B of Table 2 report the ADEs defined in (2.8) with Newey-West standard errors. As a matter of comparison, we also report the OLS estimates of the linear regression of  $\Delta f_t$  on the (standardized) predictors  $\mathbf{x}_{t-1}^*$ . The ADEs must yield similar values to the OLS estimates if the relation between  $\Delta f_t$  and  $\mathbf{x}_{t-1}^*$  is truly linear. Indeed, both estimators are supposed to measure the same thing because, when  $m(\cdot)$  is the identity function,  $\gamma$  equals unity in (2.3).

For the daily horizon, the Wald statistics reject the hypothesis that all ADEs are jointly equal to zero. This is less obvious for the OLS estimates. More importantly, with the ADE model, the lagged futures returns do always carry significant information about the next period futures returns. In contrast, the two-period lagged spot returns have no significant impact, with the possible exception of corn with the ADE model and soybeans with the OLS model. The squared futures and spot returns do not have a significant explanatory power for any

commodity and the interaction component only matters for corn. Although it is not significant, the estimated effect of the basis on the futures returns is always negative (with the exception of soybeans with ADE), as expected. For the weekly horizon, the uniformity of the results across commodities is much weaker. Moreover, the predictive power of the ADE model disappears for corn, soybeans and wheat, as indicated by the Wald tests. As can be seen from the two last columns of Panel B, there is predictability for oats on a weekly horizon. However, when plotting the response function  $\hat{E} \left[ \Delta f_t | \mathbf{x}_{t-1}^{*'} \hat{\boldsymbol{\delta}}^{IV} \right]$  (not reported here), it appears that the predictability area is smaller than the one from the daily horizons. For all these reasons, we focus on daily data for the remainder of the analysis.

The ADE and OLS estimates are used as weights to construct two daily indices,  $\mathbf{x}_{t-1}^{*'} \hat{\boldsymbol{\delta}}^{IV}$  and  $\mathbf{x}_{t-1}^{*'} \hat{\boldsymbol{\beta}}^{OLS}$ . It can be shown (not reported here) that all the ADE indices based on the full set of (standardized) predictors are mainly driven by  $(\Delta s_{t-2})^2$  and, to a lesser extent, by  $(\Delta f_{t-1})^2$  and/or  $\Delta f_{t-2} \Delta s_{t-2}$ . To get some insight, take the example of soybeans: the ADEs of  $(\Delta s_{t-2})^2$ ,  $(\Delta f_{t-1})^2$  and  $\Delta f_{t-2} \Delta s_{t-2}$  are much larger in absolute value than the coefficient estimates of  $f_{t-2} - s_{t-2}$ ,  $\Delta s_{t-2}$ ,  $\Delta f_{t-1}$ . This clearly adds noise to the indices as these standardized predictors have been shown not to be significant. To circumvent that problem and to avoid potential data snooping critiques by selecting a particular set of variables for each different commodity, we construct all indices from the constrained regression where we impose the coefficients of  $(\Delta f_{t-1})^2$ ,  $(\Delta s_{t-1})^2$  and  $\Delta f_{t-1} \Delta s_{t-1}$  to be zero: that is we use

$$\mathbf{x}_{t-1}^{**} = (\Delta f_{t-1}, \Delta s_{t-2}, f_{t-2} - s_{t-2})'.$$

ADE and OLS estimates of the constrained regression are reported in Panel C of Table 2. Estimates of the remaining coefficients are qualitatively similar to the ones reported in Panel A. In Table 3, we show the descriptive statistics of the constrained ADE and OLS indices. There is a reassuringly high correlation between the two indices.

The next step is to model the response of the expected futures returns to the estimated index  $\mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV}$ . It is worth noting that the Nadaraya-Watson kernel estimator  $\hat{m} \left( \mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV} \right)$  converges at the optimal nonparametric (univariate) rate  $T^{-2/5}$ , even though  $\mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV}$  is replaced by the estimated index  $\mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV}$  (Härdle and Stoker (1989), Theorem 3.3). In Figures 1 to 4, we plot the response functions  $\hat{E} \left[ \Delta f_t | \mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV} \right] \equiv \hat{m} \left( \mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV} \right)$ . To diminish the distorting influence of the outliers,  $\hat{m} \left( \mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV} \right)$  is estimated on the interval bounded by the sample mean of  $\mathbf{x}_{t-1}^{**'} \hat{\boldsymbol{\delta}}^{IV}$  plus and minus twice its standard deviation. A 95% bootstrap pointwise confidence

Table 2: Average derivative and OLS estimates for prediction of futures returns, 1/1/1979-31/12/2003.

DAILY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-2} - s_{t-2}$	-0.0023 [0.0327]	-0.0001 [0.0219]	-0.0079 [0.0306]	-0.0102 [0.0207]	0.0052 [0.0554]	-0.0270 [0.0192]	-0.0295 [0.0359]	-0.0450 [0.0277]
$\Delta s_{t-2}$	-0.0696 [0.0342]	0.0085 [0.0199]	-0.0186 [0.0432]	-0.0382 [0.0206]	0.0060 [0.0676]	0.0403 [0.0203]	0.1033 [0.0711]	0.0040 [0.0293]
$\Delta f_{t-1}$	0.1067 [0.0291]	0.0593 [0.0223]	0.1391 [0.0366]	0.0340 [0.0213]	-0.0906 [0.0380]	-0.0255 [0.0232]	0.1674 [0.0429]	0.0983 [0.0294]
$(\Delta f_{t-1})^2$	0.7959 [0.7830]	0.0073 [0.0280]	-0.0448 [0.3835]	-0.0003 [0.0333]	1.5929 [3.1065]	-0.0274 [0.0282]	0.0510 [0.1559]	-0.0087 [0.0321]
$(\Delta s_{t-2})^2$	1.3375 [1.3390]	-0.0557 [0.0452]	-0.3539 [1.1846]	-0.0315 [0.0252]	2.1606 [4.5095]	-0.0146 [0.0667]	0.0405 [0.5080]	0.0575 [0.0453]
$\Delta f_{t-2} \Delta s_{t-2}$	0.4007 [0.1825]	0.1017 [0.0415]	-0.0166 [0.1309]	0.0364 [0.0292]	0.4939 [0.4660]	0.0487 [0.0538]	0.2826 [0.1704]	0.0448 [0.0433]
Wald stat	21.9420	11.5755	15.4056	5.9680	13.0639	11.7784	21.6758	19.5770
<i>p</i> -val.	0.0012	0.0721	0.0173	0.4268	0.0420	0.0671	0.0014	0.0033

WEEKLY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-2} - s_{t-2}$	-0.1828 [0.1662]	0.1256 [0.1278]	0.2253 [0.1560]	0.0396 [0.0977]	-0.1599 [0.1591]	-0.1163 [0.0822]	-0.3212 [0.1741]	-0.1107 [0.1402]
$\Delta s_{t-2}$	0.1687 [0.2005]	0.2584 [0.0957]	0.2180 [0.1826]	0.0222 [0.1136]	0.1393 [0.1622]	0.2037 [0.0988]	0.1302 [0.2517]	0.1860 [0.1295]
$\Delta f_{t-1}$	-0.0005 [0.2372]	-0.0519 [0.1130]	-0.0887 [0.1807]	0.0592 [0.0919]	0.0647 [0.1404]	-0.1569 [0.1126]	-0.1498 [0.2193]	-0.2329 [0.1820]
$(\Delta f_{t-1})^2$	0.4496 [3.4684]	0.0969 [0.2231]	-1.6213 [0.7933]	-0.1743 [0.1476]	-1.9356 [2.4883]	-0.2313 [0.1393]	-0.7590 [3.0644]	0.0055 [0.3014]
$(\Delta s_{t-2})^2$	0.0873 [4.2104]	0.1877 [0.1471]	-2.6008 [1.8326]	-0.1955 [0.0705]	-4.0273 [2.9643]	-0.5643 [0.3131]	-0.8692 [4.8540]	-0.2551 [0.2407]
$\Delta f_{t-2} \Delta s_{t-2}$	0.1823 [0.6581]	-0.1096 [0.2461]	-0.7113 [0.5008]	0.2624 [0.1387]	1.2849 [0.8733]	0.8855 [0.3599]	2.4740 [0.8414]	0.5725 [0.2331]
Wald stat	3.2242	12.8650	13.2046	20.4691	7.3072	13.3083	16.6477	14.5278
<i>p</i> -val.	0.7802	0.0452	0.0599	0.0023	0.2934	0.0384	0.0107	0.0243

DAILY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-2} - s_{t-2}$	-0.01181 [0.03177]	0.00307 [0.02085]	-0.01059 [0.02919]	-0.00995 [0.02046]	-0.00376 [0.02835]	-0.02577 [0.01993]	-0.05543 [0.03358]	-0.04971 [0.02919]
$\Delta s_{t-2}$	-0.03735 [0.02661]	0.00693 [0.02096]	-0.03378 [0.03059]	-0.02849 [0.01934]	0.00985 [0.02669]	0.03855 [0.02092]	0.05313 [0.05814]	0.00101 [0.02864]
$\Delta f_{t-1}$	0.06404 [0.02403]	0.06039 [0.02290]	0.08568 [0.02906]	0.03418 [0.02144]	-0.08398 [0.02522]	-0.02439 [0.02354]	0.12370 [0.03406]	0.09913 [0.02918]
Wald stat	8.9393	7.0092	9.9362	4.3164	11.3207	7.5010	17.2495	13.498
<i>p</i> -val.	0.0301	0.0716	0.0191	0.2293	0.0101	0.0573	0.0006	0.0037

a. Newey-West standard errors in brackets and robust Wald tests.

b. All estimates and standard errors have been multiplied by  $10^2$ .

interval is also added.<sup>15</sup>

<sup>15</sup>The optimal cross-validation bandwidths are 0.0007185, 0.0005870, 0.0005670 and 0.0009252 for corn, wheat, soybeans and oats, respectively. The trimming bound is set such that 5% of the observations are dropped. We



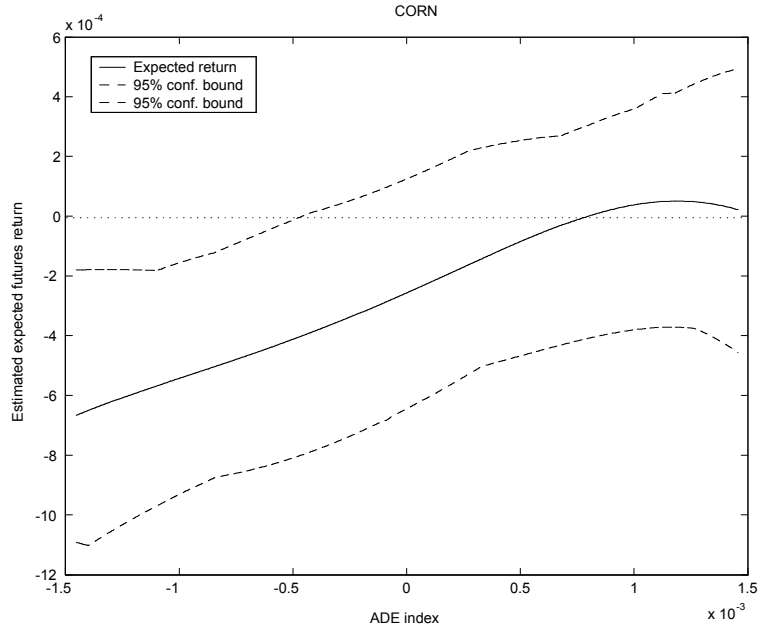


Figure 1: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{**} = (f_{t-2} - s_{t-2}, \Delta f_{t-1}, \Delta s_{t-2})'$ ; corn, 1/1/1979-31/12/2003.

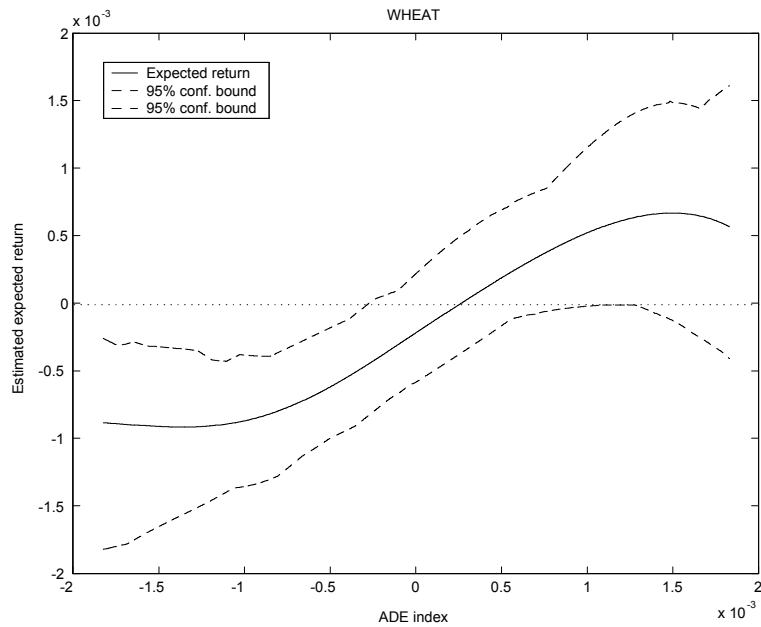


Figure 2: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{**} = (f_{t-2} - s_{t-2}, \Delta f_{t-1}, \Delta s_{t-2})'$ ; wheat, 1/1/1979-31/12/2003.

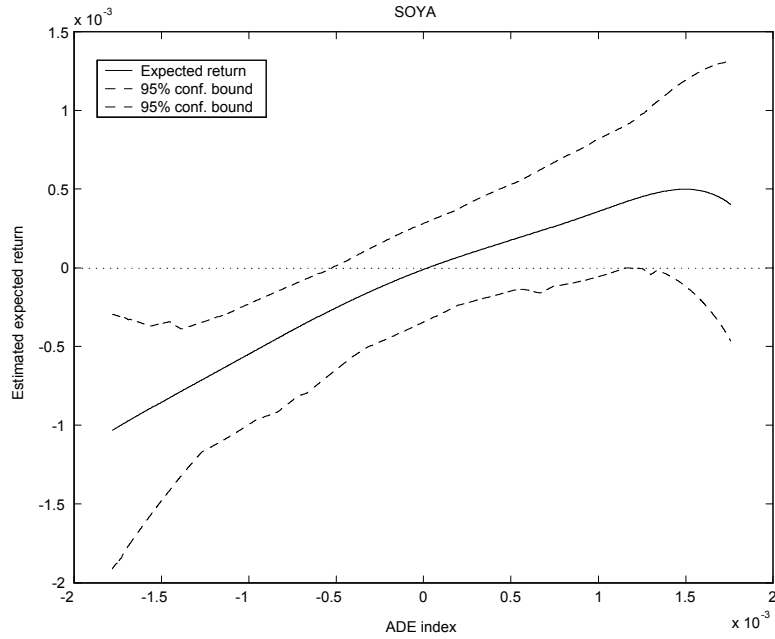


Figure 3: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{**} = (f_{t-2} - s_{t-2}, \Delta f_{t-1}, \Delta s_{t-2})'$ ; soya, 1/1/1979-31/12/2003.

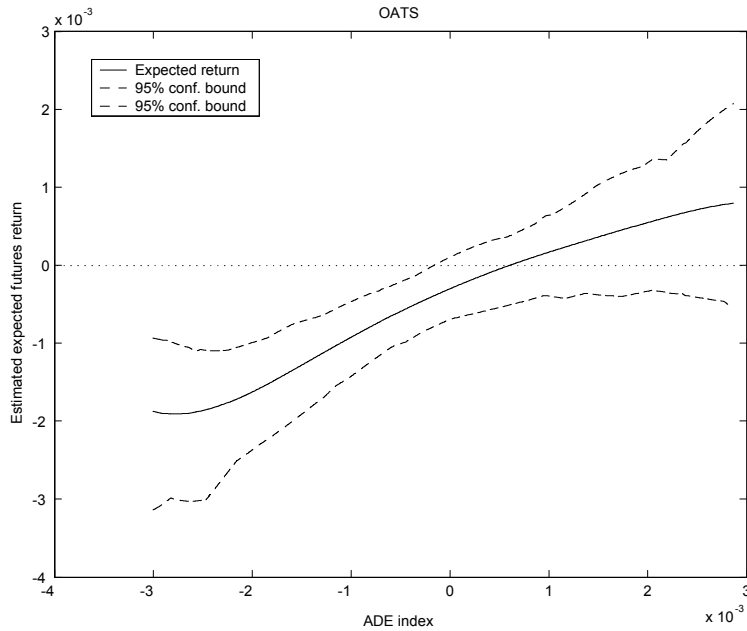


Figure 4: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{**} = (f_{t-2} - s_{t-2}, \Delta f_{t-1}, \Delta s_{t-2})'$ ; oats, 1/1/1979-31/12/2003.

Table 3: Descriptive statistics of the daily ADE- and OLS-based indices with 3 predictors, 1/1/1979-31/12/2003.

DAILY	Corn		Wheat	
	ADE	OLS	ADE	OLS
mean	$-1.54 \times 10^{-19}$	$-9.54 \times 10^{-8}$	$-3.10 \times 10^{-20}$	$-1.69 \times 10^{-8}$
var.	$7.25 \times 10^{-4}$	$3.77 \times 10^{-7}$	$9.15 \times 10^{-4}$	$2 \times 10^{-7}$
skew.	0.0468	-0.0066	0.0414	0.1087
kurt.	5.2218	5.8684	5.4424	6.9870
correl.	0.7744		0.9430	
DAILY	Soybeans		Oats	
	ADE	OLS	ADE	OLS
mean	$-3.17 \times 10^{-20}$	$-4.41 \times 10^{-8}$	$4.57 \times 10^{-8}$	$1.98 \times 10^{-7}$
var.	$8.43 \times 10^{-4}$	$2.70 \times 10^{-7}$	0.0015	$1.26 \times 10^{-6}$
skew.	0.1013	-0.1527	0.2057	0.2597
kurt.	4.9073	6.4622	5.9414	4.7963
correl.	0.5394		0.9383	

The confidence bounds confirm that there exists predictability through the index that we consider, notably when  $\mathbf{x}_{t-1}^{**'} \hat{\delta}^{IV}$  takes on negative values. For corn and wheat, this happens when the lagged futures return takes on large negative values and/or the two-period lagged spot return and basis take on large and positive values. For oats, the index is negative when the two-period lagged basis is positive and the two-period lagged spot and one-period lagged futures returns are negative. For soybeans, there is predictability when the basis and the lagged futures return are positive and the two-period lagged spot return is negative. We also observe an ‘‘asymmetry’’ in predictability: all indices are only able to predict negative futures returns. The graphs also indicate that the relation between  $\mathbf{x}_{t-1}^{**'} \hat{\delta}^{IV}$  and the expected return deviates from linearity.

Recall that all of the above ignores information that became available in the late afternoon and the subsequent morning preceding the prediction period, which is noon to noon. In that sense, what we found thus far is a lower bound on the predictability. Including the missing information is possible only at the cost of including also the early-afternoon data, which is not yet available when the prediction interval starts. Still, doing the tests with this set of data provides an upper bound on the predictability and allows us to check whether the forecasting pattern observed thus far is much affected by subsequent information. All results are reported in Tables 5 and 6 in the appendix. For the same reasons as above, we ended with the three

---

use 100 bootstrap replications and choose an average block length of 10 periods. This value seems reasonable given the weak autocorrelation in the daily data.

following predictors:

$$\mathbf{x}_{t-1}^{***} = (\Delta f_{t-1}, \Delta s_{t-1}, f_{t-1} - s_{t-1})'.$$

The ADEs are presented in Panel C of Table 5. The spot return predictor becomes significant, which is not surprising when the period covered by the spot return overlaps with that covered by the futures return. Although not significant, the sign of the basis is still negative.

The curves  $\hat{m}(\mathbf{x}_{t-1}^{***} \hat{\delta}^{IV})$  are graphed in Figures 5 to 8.<sup>16</sup> In order to compare both approaches, we focus on the area where there is predictability, notably the area where the index assumes negative values. Notice that, for all commodities, the upper confidence bounds are below the zero line. Thus, here, a way to compare both approaches with respect to  $H_{0,\mathbf{x}}$  is to compare the distance from the 95% upper bounds to the zero line. Indeed, the further away an upper bound is from the zero line, the stronger the rejection is of the null  $H_{0,\mathbf{x}}$  and vice versa. For corn, soybeans and oats in Figure 5, 7 and 8, the upper confidence bound are generally further away from zero line than in Figure 1, 3, and 4, indicating a stronger rejection of  $H_{0,\mathbf{x}}$ ; only in Figure 6 (wheat) is the upper bound closer than its analog in Figure 2, suggesting a weaker rejection of  $H_{0,\mathbf{x}}$ . In short, the extra information would have allowed the trader to make more pronounced negative forecasts, but would not have affected the near-zero forecasts. This confirms the asymmetry we established before: there is no indication that the information missed in the base test would have allowed to predict also positive returns.

### 3.3 Economic implications

In this section we analyze the consequences of the observed predictability for the hedger. Recall that, for the hedger, our question is whether any predictability has a noticeable impact on the optimal hedge ratio. For illustrative purposes, we first compute the hedge ratio at the one-day horizon for the pure hedging model, the static ratio  $\hat{h}^{MinVar} = \frac{\widehat{\text{cov}}(\Delta s_t, \Delta f_t)}{\widehat{\text{var}}(\Delta f_t)}$ . Next we compute the differential impact of the predictability on the hedge ratio, for values of  $A$  between two and ten and for a given value of the ADE index, set at the sample mean of the ADE index minus one standard deviation:

$$\Delta \hat{h}_{t-1} = -\frac{\frac{1}{A} \hat{E} [\Delta f_t | \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]}{\widehat{\text{var}}(\Delta f_t)}.$$

---

<sup>16</sup>The optimal cross-validation bandwidths are 0.0012498, 0.0006867, 0.001044 and 0.0014072 for corn, wheat, soybeans and oats, respectively. The trimming bound stays the same.

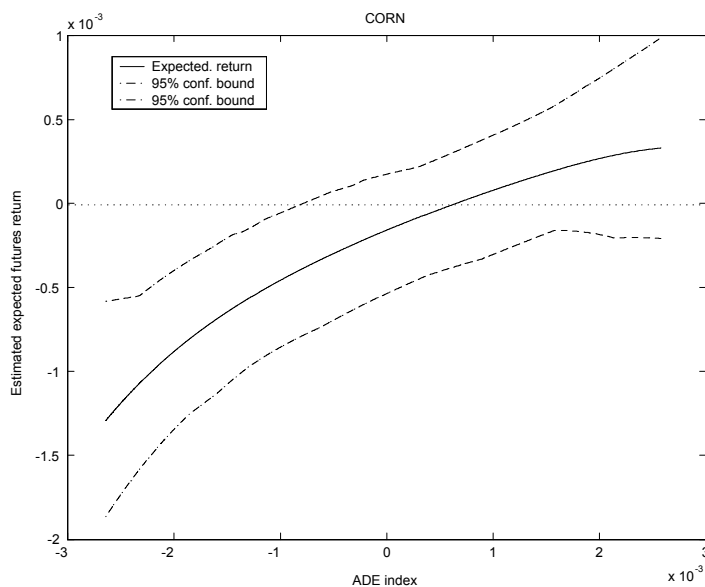


Figure 5: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{***} = (f_{t-1} - s_{t-1}, \Delta f_{t-1}, \Delta s_{t-1})'$ ; corn, 1/1/1979-31/12/2003.

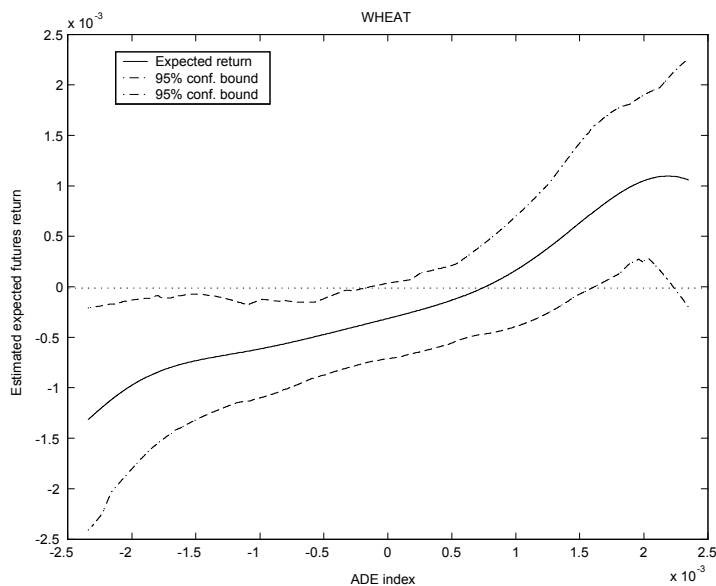


Figure 6: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{***} = (f_{t-1} - s_{t-1}, \Delta f_{t-1}, \Delta s_{t-1})'$ ; wheat, 1/1/1979-31/12/2003

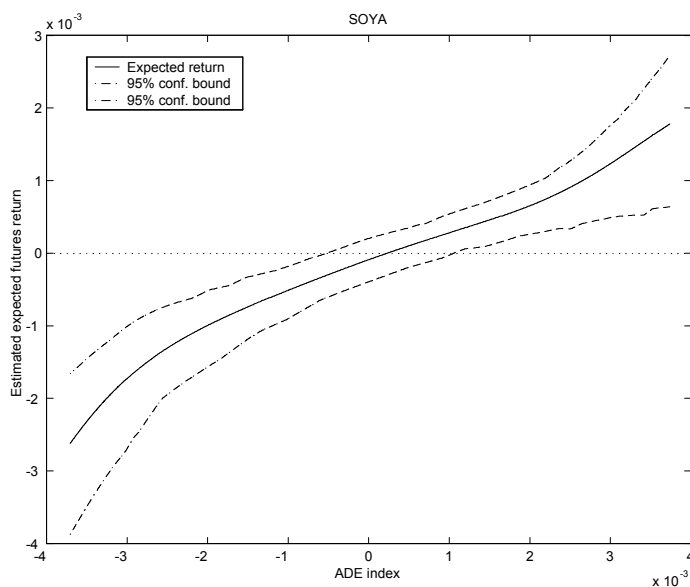


Figure 7: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{***} = (f_{t-1} - s_{t-1}, \Delta f_{t-1}, \Delta s_{t-1})'$ ; soya, 1/1/1979-31/12/2003

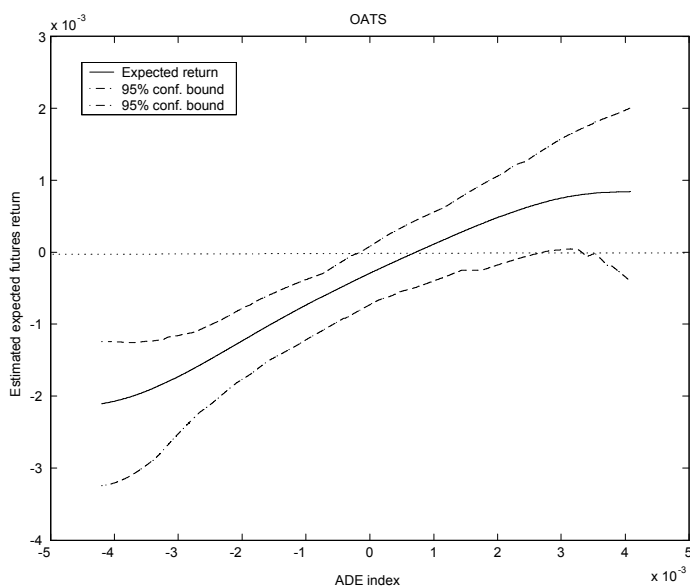


Figure 8: Estimated conditional expected return in function of the ADE index constructed with 3 predictors  $\mathbf{x}_{t-1}^{***} = (f_{t-1} - s_{t-1}, \Delta f_{t-1}, \Delta s_{t-1})'$ ; oats, 1/1/1979-31/12/2003

Table 4: Comparison of mean-variance and minimum-variance hedge ratios, 1/1/1979-31/12/2003.

		$\hat{h}^{MinVar} = \frac{\widehat{cov}(\Delta s_t, \Delta f_t)}{\widehat{var}(\Delta f_t)}$	$\Delta \hat{h}_{t-1} = -\frac{\frac{1}{A} \hat{E}[\Delta f_t]^{UB}}{\widehat{var}(\Delta f_t)}$			
$\hat{E}[\Delta f_t   \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]^{UB}$		$A = \infty$	$A = 10$	$A = 5$	$A = 3.5$	$A = 2$
Corn	$\approx -0.010\%$	0.98	0.07	0.13	0.19	0.34
Wheat	$\approx -0.025\%$	0.92	0.13	0.27	0.38	0.67
Oats	$\approx -0.025\%$	0.46	0.08	0.16	0.23	0.39
Soybeans	$\approx -0.090\%$	0.93	0.54	1.08	1.55	2.70
$\hat{E}[\Delta f_t   \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]$		$A = \infty$	$A = 10$	$A = 5$	$A = 3.5$	$A = 2$
Corn	$\approx -0.050\%$	0.98	0.34	0.67	0.96	1.68
Wheat	$\approx -0.100\%$	0.92	0.53	1.07	1.52	2.66
Oats	$\approx -0.075\%$	0.46	0.24	0.47	0.68	1.18
Soybeans	$\approx -0.150\%$	0.93	0.90	1.80	2.58	4.51
$\hat{E}[\Delta f_t   \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]^{LB}$		$A = \infty$	$A = 10$	$A = 5$	$A = 3.5$	$A = 2$
Corn	$\approx -0.090\%$	0.98	0.60	1.21	1.73	3.02
Wheat	$\approx -0.150\%$	0.92	0.80	1.60	2.28	4.00
Oats	$\approx -0.100\%$	0.46	0.32	0.63	0.90	1.58
Soybeans	$\approx -0.250\%$	0.93	1.50	3.00	4.29	7.51

Note:  $\hat{E}[\Delta f_t | \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]^{LB (UB)}$  is the 95% confidence lower (upper) bound of the estimated expected return given that the index equals its sample mean minus once its standard deviation.  $\hat{E}[\Delta f_t | \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]$  is the estimated expected return given that the index equals its sample mean minus once its standard deviation.

Results are reported in Table 4 for all commodities. As expected, the static hedge ratios  $\hat{h}^{MinVar}$  are close to unity, with the exception of oats. The reason might be that the oats spot market is the Minneapolis exchange whereas the futures contract is traded on the Chicago Board of Trade. For a degree of risk of aversion  $A = 3.5$ , the impact of the mean on  $\hat{h}_{t-1}^{\mu\sigma^2}$  is potentially huge, even when using the 95% upper bound of the estimated futures return  $\hat{E}[\Delta f_t | \mathbf{ADE} = \overline{\mathbf{ADE}} - \hat{\sigma}(\mathbf{ADE})]^{UB}$  instead of the central forecast. But even for  $A = 10$ , central-forecast hedge ratios still rise by at least 30 percent, and occasionally even almost 100 percent. It is also clear that the large 95% confidence interval around the estimated expected mean, has a high impact on the mean-variance hedge ratios, but never to the extent that it would make the minimum-variance hedge a serious candidate for the optimal ratio. In short, we can conclude there is an economically important tradeoff between risk and return in hedging, for a certain range of negative values of the index, even though the mean-variance hedge ratio remains imprecisely determined due to the lack of precision in the predictability.

For the academic economist, the puzzle is why any predictability is confined to negative forecasts. It is hard to believe that the commodity futures' non-diversifiable risk, its covariance with the pricing kernel, goes from zero to negative on a day to day basis; in fact, estimated

betas for commodity futures have turned positive, recently.<sup>17</sup> There seems to be an unidentified cost in going short.

## 4 Conclusion

In this paper, we investigated the predictability of the agricultural futures returns and its practical implications. The question was addressed via a semiparametric approach where a parametric index combining several predictors is nonparametrically related to the expected futures returns.

We find statistically strong empirical evidence against Samuelson's (1965) hypothesis. Our results indicate that each estimated index contains statistically significant information regarding the expected futures returns. This finding has a large effect on the optimal mean-variance hedge ratio, even though it is still estimated rather imprecisely.

---

<sup>17</sup>The Economist, Oct 12, 2006.



## Appendix: Additional results

Table 5: Average derivative and OLS estimates for prediction of futures returns, 1/1/1979-31/12/2003.

DAILY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-1} - s_{t-1}$	-0.0352	-0.0044	-0.0327	-0.0124	-0.0278	-0.0400	-0.0369	-0.0574
	[0.0316]	[0.0205]	[0.0317]	[0.0207]	[0.0296]	[0.0210]	[0.0354]	[0.0284]
$\Delta s_{t-1}$	0.2048	0.1561	0.1617	0.0506	0.3542	0.1880	0.1410	0.0511
	[0.0532]	[0.0337]	[0.1004]	[0.0328]	[0.0717]	[0.0464]	[0.0736]	[0.0303]
$\Delta f_{t-1}$	-0.1023	-0.0691	-0.0424	-0.0029	-0.3847	-0.1899	0.1590	0.0837
	[0.0508]	[0.0338]	[0.0950]	[0.0321]	[0.0673]	[0.0454]	[0.0487]	[0.0314]
$(\Delta f_{t-1})^2$	0.4237	0.0312	0.0451	-0.0188	-0.8803	0.0756	0.2273	-0.0075
	[0.7973]	[0.0731]	[3.8606]	[0.0592]	[0.8956]	[0.0840]	[0.4874]	[0.0355]
$(\Delta s_{t-1})^2$	0.6959	-0.0098	0.8701	0.0113	-0.9877	-0.0014	1.0112	0.0022
	[1.1491]	[0.0389]	[9.7229]	[0.0241]	[1.0810]	[0.0551]	[1.2205]	[0.0374]
$\Delta f_{t-1} \Delta s_{t-1}$	-0.6078	0.0007	-0.0460	0.0273	1.3399	-0.1091	-0.2350	0.0160
	[1.1443]	[0.0895]	[5.1060]	[0.0572]	[1.3850]	[0.1007]	[0.3540]	[0.0446]
Wald stat	23.7490	33.3857	19.7401	5.1510	34.2444	27.2121	25.6076	17.1708
<i>p</i> -val.	0.0006	0.0000	0.0031	0.5246	0.0000	0.0001	0.0003	0.0087

WEEKLY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-1} - s_{t-1}$	-0.1543	-0.0033	-0.0287	0.0165	-0.0094	-0.0326	-0.3800	-0.1749
	[0.1507]	[0.0986]	[0.1467]	[0.0877]	[0.1434]	[0.0930]	[0.1882]	[0.1685]
$\Delta s_{t-1}$	-0.2419	0.2083	-0.0888	-0.0347	0.0850	-0.4199	0.1611	0.1768
	[0.3349]	[0.2059]	[0.3260]	[0.2136]	[0.3011]	[0.2405]	[0.2761]	[0.2121]
$\Delta f_{t-1}$	0.1789	-0.2613	0.0448	0.0573	0.0250	0.2509	-0.2746	-0.3132
	[0.3237]	[0.2013]	[0.3103]	[0.1907]	[0.2974]	[0.2348]	[0.3018]	[0.2464]
$(\Delta f_{t-1})^2$	1.8086	0.3748	-1.0830	-0.4214	-0.1301	1.1520	-2.6250	-0.2624
	[2.5574]	[0.4687]	[1.8978]	[0.3191]	[1.9616]	[0.7713]	[7.2469]	[0.3834]
$(\Delta s_{t-1})^2$	0.8783	-0.2568	-0.0141	0.2305	0.1405	0.0063	-0.9773	0.1004
	[2.7560]	[0.1380]	[3.2574]	[0.1478]	[1.8111]	[0.4503]	[8.1461]	[0.1451]
$\Delta f_{t-1} \Delta s_{t-1}$	-2.4121	-0.0543	0.2330	0.2083	0.2643	-1.3097	1.5323	0.3587
	[3.2103]	[0.4138]	[2.3315]	[0.3356]	[2.9735]	[1.0325]	[5.6145]	[0.3145]
Wald stat	3.5738	20.4350	4.5205	4.8219	0.8464	7.3869	11.1592	7.6810
<i>p</i> -val.	0.7341	0.0023	0.6066	0.5669	0.9908	0.2865	0.0836	0.2624

DAILY	Corn <sup>a,b</sup>		Wheat <sup>a,b</sup>		Soybeans <sup>a,b</sup>		Oats <sup>a,b</sup>	
	ADE	OLS	ADE	OLS	ADE	OLS	ADE	OLS
$f_{t-1} - s_{t-1}$	-0.0400	-0.0039	-0.0280	-0.0108	-0.0254	-0.0454	-0.0498	-0.0576
	[0.0310]	[0.0197]	[0.0278]	[0.0203]	[0.0288]	[0.0224]	[0.0342]	[0.0295]
$\Delta s_{t-1}$	0.2150	0.1563	0.1470	0.0531	0.3434	0.2133	0.1033	0.0507
	[0.0459]	[0.0315]	[0.0606]	[0.0324]	[0.0632]	[0.0554]	[0.0622]	[0.0303]
$\Delta f_{t-1}$	-0.1345	-0.0694	-0.0448	-0.0056	-0.3638	-0.2112	0.1422	0.0835
	[0.0430]	[0.0333]	[0.0514]	[0.0287]	[0.0607]	[0.0524]	[0.0407]	[0.0311]
Wald stat	26.2838	30.2952	11.0370	4.4634	37.1776	19.9870	28.4503	14.4058
<i>p</i> -val.	0.0000	0.0000	0.0115	0.2156	0.0000	0.0002	0.0000	0.0024

a. Newey-West standard errors in brackets and robust Wald tests.

b. All estimates and standard errors have been multiplied by  $10^2$ .

Table 6: Descriptive statistics of the daily ADE- and OLS-based indices with 3 predictors, 1/1/1979-31/12/2003.

DAILY	Corn		Wheat	
	ADE	OLS	ADE	OLS
mean	$-4.91 \times 10^{-19}$	$-3.23 \times 10^{-8}$	$1.03 \times 10^{-19}$	$-1.48 \times 10^{-7}$
var.	$1.78 \times 10^{-6}$	$1.12 \times 10^{-6}$	$1.51 \times 10^{-6}$	$0.26 \times 10^{-6}$
skew.	-0.8557	-0.8688	-2.2868	-1.4390
kurt.	20.8220	16.4510	48.8536	28.6041
correl.	0.9327		0.9842	
	Soybeans		Oats	
	ADE	OLS	ADE	OLS
mean	$-2.33 \times 10^{-19}$	$-9.37 \times 10^{-8}$	$3.66 \times 10^{-18}$	$2.03 \times 10^{-7}$
var.	$2.95 \times 10^{-6}$	$1.30 \times 10^{-6}$	$4.53 \times 10^{-6}$	$1.60 \times 10^{-6}$
skew.	0.1623	0.0209	0.0659	0.2601
kurt.	27.8915	25.3289	7.0035	5.9467
correl.	0.9580		0.9723	

## References

- [1] Baillie, R. T., Myers, R. J. (1991). Bivariate estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, 6, 109-124.
- [2] Bera, A. K., Garcia, P., Roh, J.-S. (1997). Estimation of time-varying hedge ratios for corn and soybeans: BGARCH and random coefficient approaches. *The Indian Journal of Statistics*, 59, 346-368.
- [3] Bessembinder, H., Chan, K. (1992). Time-varying risk premia and forecastable returns in futures markets, *Journal of Financial Economics*, 32, 169-193.
- [4] Byström, H. N. E. (2003). The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics*, 35, 1-11.
- [5] Campbell, J. Y., Lo, A. W., MacKinley, A. C. (1996). *The econometrics of financial markets*. Princeton University Press.
- [6] Chen, X., Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 66, 289-314.
- [7] Ederington, L. H. (1979). The hedging performance of the new futures markets. *Journal of Finance*, 34, 157-170.
- [8] Ghysels, E., Ng, S. (1998). A semiparametric factor model of interest rates and tests of the affine term structure. *Review of Economics and Statistics*, 80, 535-548.

- [9] Härdle, W. (1990). Applied nonparametric regression, Cambridge University Press.
- [10] Härdle, W., Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84, 986-995.
- [11] Harris, R. D. F., Shen, J. (2003). Robust estimation of the optimal hedge ratio. *Journal of Futures Markets*, 23, 799-816.
- [12] Hayashi, F. (2000). Econometrics. Princeton University Press.
- [13] Hsin, C. W., Kuo, J., Lee, C. F. (1994). A new measure to compare the hedging effectiveness of foreign currency futures. *Journal of Futures Markets*, 14, 685-707.
- [14] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics*, 58, 71-120
- [15] Ichimura, H., Todd, P. E. (2006). Implementing parametric and nonparametric estimators. To appear in the Handbook of Econometrics, Volume 5. See <http://athena.sas.upenn.edu/~petra/wbank/curver9.pdf>
- [16] Johnson, L. L. (1960). The theory of hedging and speculation in commodity futures. *Review of Economic Studies*, 27, 139-151.
- [17] LeRoy, S. F. (1989). efficient capital Markets and Martingales. *Journal of Economic Literature*, 27, 1583-1621.
- [18] Li, W. (1996). Asymptotic equivalence of average derivatives. *Economics Letters*, 52, 241-245.
- [19] Lien, D.- H. D., Tse, Y. K., Tsui, A. K. (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Economics*, 12, 791-798.
- [20] McCurdy, T. H., Morgan, I. G . (1988). Testing the Martingale Hypothesis in Deutsche Mark Futures with Models Specifying the Form of Heteroscedasticity. *Journal of Applied Econometrics*, 3, 187-202.
- [21] Miffre, J. (2002). Economic significance of the predictable movements in futures returns. *Economic Notes*, Economic Notes, 31, 125-142.
- [22] Pagan, A., Ullah, A. (1999). Nonparametric Econometrics, Cambridge University Press.

- [23] Politis, D. N., Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303-1313.
- [24] Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6, 41-49.
- [25] Silverman, B. (1986). Density estimation for statistics and data analysis. Chapman and Hall: London.
- [26] Stein, J. L. (1961). The simultaneous determination of spot and futures prices. *American Economic Review*, 51, 1012-1026.
- [27] Stoker, T. (1986). Consistent estimator of scaled coefficients. *Econometrica*, 54, 1461-1481.
- [28] Stoker, T. (1993). Smoothing bias in density derivative estimation. *Journal of the American Statistical Association*, 88, 855-863.