Journal of Financial Economics 67 (2003) 473–509
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/econbase

The economic value of volatility timing using "realized" volatility ☆

Jeff Fleming[a,*], Chris Kirby[b], Barbara Ostdiek[a]

[a] *Jones Graduate School, Rice University, Houston, TX 77005, USA*
[b] *School of Management, University of Texas at Dallas, Richardson, TX 75083, USA*

Received 20 November 2001; accepted 1 February 2002

Abstract

Recent work suggests that intradaily returns can be used to construct estimates of daily return volatility that are more precise than those constructed using daily returns. We measure the economic value of this "realized" volatility approach in the context of investment decisions. Our results indicate that the value of switching from daily to intradaily returns to estimate the conditional covariance matrix can be substantial. We estimate that a risk-averse investor would be willing to pay 50 to 200 basis points per year to capture the observed gains in portfolio performance. Moreover, these gains are robust to transaction costs, estimation risk regarding expected returns, and the performance measurement horizon.

JEL classification: G11; G14

Keywords: Volatility timing; Realized volatility; Portfolio optimization; Mean-variance analysis; Rolling estimators

☆We thank Tim Bollerslev for providing useful comments during the initial stages of this research. We also thank the referee and Eric Ghysels and Mike Stutzer for comments on an earlier draft as well as seminar participants at the 2001 AGSM Accounting and Finance Research Camp, the University of Alabama, Georgetown University, Georgia State University, the University of Houston, the University of South Carolina, the University of Texas at Dallas, and Washington University.

*Corresponding author.
E-mail address: jfleming@rice.edu (J. Fleming).

0304-405X/02/$ - see front matter
PII: S0304-405X(02)00259-3

1. Introduction

Volatility modeling is important for many option pricing, portfolio selection, and risk management applications.[1] A number of volatility models have been developed over the years, including ARCH (Engle, 1982), GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), and stochastic volatility specifications (see Taylor, 1986), and the performance of these models has been evaluated exhaustively. Recently, however, the volatility modeling literature appears to take a significant step forward. Andersen, Bollerslev, Diebold and Labys (2001) (ABDL, hereafter) and Barndorff-Nielsen and Shephard (2002), building on earlier work by Schwert (1989) and Hsieh (1991), propose a new approach called "realized" volatility that exploits the information in high-frequency returns. Basically, the approach is to estimate volatility by summing the squares of intradaily returns sampled at very short intervals. The idea is that if the sample path of volatility is continuous, then increasing the sampling frequency yields arbitrarily precise estimates of volatility at any given point in time (Merton, 1980).[2] In effect, volatility becomes observable.

A number of papers implement this approach and examine the properties of realized volatility. ABDL (2001) examine currencies, Andersen, Bollerslev, Diebold and Ebens (2001) (ABDE, hereafter) examine individual stocks, Ebens (1999) examines the Dow Jones Industrial Average, and Areal and Taylor (2002) examine stock index futures. The results are generally consistent across assets and are quite compelling. For example, consistent with Clark (1973), realized volatility appears to be lognormally distributed and daily returns standardized by realized volatility are approximately normal. In addition, realized volatility exhibits long-memory dynamics consistent with a fractionally integrated process with a degree of integration around 0.4, volatility clustering is apparent at as long as the monthly level, and realized volatility obeys precise scaling laws under temporal aggregation.

Although these findings provide support for the realized volatility approach, they are mainly statistical in nature. A separate question is whether the gains in precision are sufficient to have a meaningful impact on decisions that depend on conditional volatility estimates. Presumably, applications such as risk management should benefit because performance in this context depends largely on the statistical properties of the estimates. It is not clear, however, whether using realized volatility leads to more accurate option prices or better investment management decisions. Perhaps standard volatility models provide a sufficient representation of volatility dynamics for these purposes so that switching to realized volatility yields only small benefits.

This paper evaluates the economic benefits of the realized volatility approach in the context of investment decisions. Our analysis builds on the framework developed

[1] We use the term "volatility" generically to refer to any element of the covariance matrix of asset returns. Our terminology is more precise when the distinction between standard deviations, variances, covariances, and correlations is important.

[2] In practice, the benefits of more frequent sampling may be reduced by dependence and nonnormality in returns induced by noncontinuous and irregular price observations, discrete prices, and bid-ask bounce. Bai et al. (2001) provide an analysis of this issue.

in Fleming et al. (2001) (FKO, hereafter). We consider a risk-averse investor who uses conditional mean-variance analysis to allocate funds across four asset classes: stocks, bonds, gold, and cash. The investor rebalances his portfolio daily but treats expected returns as constant since there is little evidence that changes in expected returns are detectable at the daily level. This implies that the investor follows a volatility-timing strategy where his portfolio weights vary only with changes in his estimates of the conditional covariance matrix of daily returns. What we evaluate is whether using the realized volatility approach to form these estimates improves the performance of the volatility-timing strategy.

We construct our estimates of the conditional covariance matrix using rolling estimators of the form analyzed by Foster and Nelson (1996) and Andreou and Ghysels (2002). One set of estimates is based on the realized variances and covariances obtained using five-minute returns. We adjust these estimates for biases induced by non-trading periods and microstructure effects relative to a second set of rolling estimates based on daily returns. We then use these two sets of estimates to isolate the economic value of the realized volatility approach. Specifically, we implement volatility-timing strategies using both sets of estimates and we evaluate the performance gains associated with switching to the realized-volatility-based estimator.

Our results indicate that the economic value of the realized volatility approach is substantial. We estimate that an investor implementing the volatility-timing strategies would be willing to pay on the order of 50 to 200 basis points per year to capture the incremental gains generated by the realized-volatility-based estimator. This represents approximately half of the overall gains to volatility timing, measured relative to the performance of an ex ante efficient static portfolio. Moreover, we find that volatility timing at the daily level leads to performance gains over longer horizons. The gains observed at the daily level are comparable to those observed using performance measurement horizons as long as a year.

The remainder of the paper is organized as follows. Section 2 describes our volatility-timing methodology and performance measurement criteria. Section 3 develops our methodology for constructing the conditional covariance matrix estimates. Section 4 describes the data and presents the preliminary empirical analysis. Section 5 evaluates the performance of the volatility-timing strategies. Section 6 examines the robustness of our results. Section 7 concludes.

2. Empirical framework

In this section, we describe our methodology for measuring the economic value of volatility timing using realized volatility. Our approach, which follows FKO (2001), is to evaluate the empirical performance of conditionally mean-variance efficient portfolios that are rebalanced daily based on estimates of the conditional covariance matrix of daily returns. FKO (2001) form their estimates of this matrix using daily returns. We treat these results as a benchmark and focus on the incremental value of employing more precise estimates of the covariance matrix constructed using the realized volatility approach of ABDL (2001).

2.1. Volatility-timing strategies

We consider a risk-averse investor who allocates funds across a set of $N+1$ securities: N risky assets plus cash. The investor uses conditional mean-variance analysis to make his allocation decisions and rebalances his portfolio daily. To avoid restrictions on short selling and minimize transaction costs, the investor implements his allocation decisions by trading futures contracts on the risky assets.

Let $\boldsymbol{R}_t$ denote the $N \times 1$ vector of returns formed by dividing the price change for each futures contract on day t by its price on day $t-1$. Define $\mathscr{I}_{t-1}$ to be the day $t-1$ information set. To minimize conditional volatility subject to a given expected return, the investor applies the risky asset weights

$$\boldsymbol{w}_t = \frac{\mu_p \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t}{\boldsymbol{\mu}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t}, \tag{1}$$

where μ_p is the target expected return on his portfolio, $\boldsymbol{\mu}_t \equiv \mathrm{E}[\boldsymbol{R}_t | \mathscr{I}_{t-1}]$ is the $N \times 1$ vector of conditional means, and $\boldsymbol{\Sigma}_t \equiv \mathrm{E}[(\boldsymbol{R}_t - \boldsymbol{\mu}_t)(\boldsymbol{R}_t - \boldsymbol{\mu}_t)' | \mathscr{I}_{t-1}]$ is the $N \times N$ conditional covariance matrix. These weights are obtained by solving the standard portfolio optimization problem with cash playing the role of the risk-free security. The weight in cash is given by one minus the sum of the elements of $\boldsymbol{w}_t$.

Since futures do not require an initial outlay of funds, Eq. (1) indicates the number of contracts to hold for a given notional investment. The return on this portfolio should closely approximate the excess return generated by a similar portfolio invested in the underlying spot assets. Under the cost-of-carry relation, the futures return equals the spot return less the opportunity cost of funds. Thus, if this relation holds, we can replicate the return on the spot portfolio by applying the portfolio weights to the futures contracts and investing 100% of our funds in the risk-free security.

Eq. (1) implies that the optimal portfolio weights vary through time as $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ change. There is little evidence, however, that we can detect changes in expected returns at the daily level. Therefore, we assume that the investor treats $\boldsymbol{\mu}_t$ as constant and follows a volatility-timing strategy where the weights vary only with changes in $\boldsymbol{\Sigma}_t$. We consider two such strategies: one that minimizes conditional volatility subject to a target expected return (the minimum volatility strategy), and a second that maximizes expected return subject to a target conditional volatility (the maximum return strategy).

By comparing the performance of the volatility-timing strategies to that of the unconditionally efficient static portfolios with the same expected return and volatility, we can directly measure the economic value of modeling the dynamics of $\boldsymbol{\Sigma}_t$. FKO (2001) conduct this comparison using rolling estimates of $\boldsymbol{\Sigma}_t$ based on daily returns. We use these results as a benchmark and assess the incremental value of employing rolling estimates of $\boldsymbol{\Sigma}_t$ based on realized volatility. Presumably, using more precise estimates of the conditional covariance matrix will improve the performance of the volatility-timing strategies.

2.2. Measuring the value of the performance gains

Following FKO (2001), we use a utility-based approach to measure the value of the performance gains associated with using a given estimator of the conditional covariance matrix. We consider an investor with quadratic utility who each day places some fixed amount of wealth, W_0, in the risk-free asset and implements a volatility-timing strategy using futures contracts with the same notional value. The realized daily utility generated by this portfolio is

$$U(R_{pt}) = W_0\left((1 + R_f + R_{pt}) - \frac{\gamma}{2(1+\gamma)}(1 + R_f + R_{pt})^2\right), \tag{2}$$

where $R_{pt} = \boldsymbol{w}_t'\boldsymbol{R}_t$ is the portfolio return, γ is the investor's relative risk aversion, and R_f denotes the risk-free interest rate.[3]

Let R_{p_1t} and R_{p_2t} denote the returns on the volatility-timing strategies using two different estimators of the conditional covariance matrix. To measure the incremental value of using the second estimator instead of the first, we find a constant, Δ, such that $\sum_{t=1}^{T} U(R_{p_1t}) = \sum_{t=1}^{T} U(R_{p_2t} - \Delta)$. This constant represents the maximum return the investor would be willing to sacrifice each day in order to capture the performance gains associated with switching to the second estimator. We report the value of Δ as an annualized basis point fee, for two levels of relative risk aversion: $\gamma = 1$ and 10.

2.3. Assessing performance over longer horizons

Our utility-based approach for measuring value is based on the distributional properties of the volatility-timing strategies' daily returns. We focus on the daily horizon because, for a given sample period, we can estimate the variances of daily returns more precisely than those of weekly, monthly, or quarterly returns (Merton, 1980). This makes it easier to determine whether the observed performance differences are significant. More generally, however, we are interested in whether volatility timing at the daily level leads to better performance over longer horizons. The cumulative gains to volatility timing, of course, will depend on the time-series properties of the daily portfolio returns. A simple example illustrates the issue.

Let $\mu_{p(j)}$ and $\sigma^2_{p(j)}$ denote the mean and variance of the j-day portfolio return. If we invest one dollar in the portfolio on day t, then the value of our investment after j days is $\prod_{i=1}^{j} (1 + R_{pt+i})$. Thus, we can express $\mu_{p(j)}$ and $\sigma^2_{p(j)}$ as

$$\mu_{p(j)} = \mathrm{E}\left[\exp\left(\sum_{i=1}^{j} r_{pt+i}\right)\right] - 1 \tag{3}$$

and

$$\sigma^2_{p(j)} = \mathrm{E}\left[\exp\left(2\sum_{i=1}^{j} r_{pt+i}\right)\right] - \mathrm{E}\left[\exp\left(\sum_{i=1}^{j} r_{pt+i}\right)\right]^2, \tag{4}$$

[3] We use a risk-free rate of 6% in our empirical analysis.

where $r_{pt+i} = \log(1 + R_{pt+i})$. Eqs. (3) and (4) suggest that the performance of a portfolio over a multi-day horizon is determined by the mean, variance, and autocorrelation function of the continuously compounded daily portfolio returns.

We can see this more clearly by evaluating how the Sharpe ratio for the portfolio varies with the performance measurement horizon. To illustrate, consider a simple scenario in which the daily portfolio returns are randomly drawn from a univariate normal distribution. Using the normal moment generating function, we have

$$\mu_{p(j)} = (1 + \mu_{p(1)})^j - 1 \tag{5}$$

and

$$\sigma^2_{p(j)} = (1 + \mu_{p(1)})^{2j}(\exp(j\varsigma^2_{p(1)}) - 1), \tag{6}$$

where $\varsigma^2_{p(1)} \equiv \mathrm{var}(r_{pt})$. Therefore, we can express the annualized Sharpe ratio for a j-day horizon as

$$\lambda_{p(j)} = (252/j)^{1/2}\left(\frac{(1 + \mu_{p(1)})^j - 1}{(1 + \mu_{p(1)})^j(\exp(j\varsigma^2_{p(1)}) - 1)^{1/2}}\right), \tag{7}$$

where we assume there are 252 trading days per year.

Even in this simple situation it is not immediately clear how the Sharpe ratio changes with the horizon. However, we can verify numerically that for reasonable choices of the mean and volatility of r_{pt} the Sharpe ratio falls slowly with increasing j. For instance, if we take the mean and volatility of r_{pt} to be 0.04% and 1.0% per day, we obtain $\lambda_{p(1)} = 0.714$ and $\lambda_{p(252)} = 0.671$. This suggests that Sharpe ratios computed for a daily horizon should be a good guide to the performance of the volatility-timing strategies over longer horizons. An important caveat, however, is that this assumes r_{pt} is serially uncorrelated. To the extent that this assumption is violated, the actual results may differ from those suggested by our example. In Section 5 we provide empirical evidence on how the gains to volatility-timing vary with the performance measurement horizon.

3. Econometric methodology

In this section, we describe our methodology for estimating the conditional covariance matrix of daily returns. The idea, which is motivated by the work of Foster and Nelson (1996) and Andreou and Ghysels (2002), is to construct rolling estimators based on lagged returns. This approach has some significant advantages in our application. Specifically, it avoids parametric assumptions, it nests a variety of GARCH and stochastic volatility models as special cases, it is computationally efficient, and it provides a natural way to evaluate the economic value of volatility timing using realized volatility.

3.1. Rolling estimators based on daily returns

We are interested in estimating the conditional covariance matrix, $\boldsymbol{\Sigma}_t$, which we assume evolves according to some unknown time-series process. Following Foster and Nelson (1996), we construct daily estimates of this matrix using a backward-looking rolling estimator. The estimator is of the general form

$$\hat{\boldsymbol{\Sigma}}_t = \sum_{k=1}^{\infty} \boldsymbol{\Omega}_{t-k} \odot \boldsymbol{e}_{t-k}\boldsymbol{e}'_{t-k}, \tag{8}$$

where $\boldsymbol{\Omega}_{t-k}$ is a symmetric $N \times N$ matrix of weights, $\boldsymbol{e}_{t-k} = (\boldsymbol{R}_{t-k} - \boldsymbol{\mu})$ is an $N \times 1$ vector of daily return innovations, and $\odot$ denotes element-by-element multiplication. The premise behind this approach is straightforward. If $\boldsymbol{\Sigma}_t$ is time varying, then its dynamics are reflected in the sample path of daily returns. Therefore, by applying a suitable set of weights to the squares and cross products of the lagged return innovations, we can construct time-series estimates of $\boldsymbol{\Sigma}_t$.

The estimator in Eq. (8) is nonparametric in nature. For certain choices of $\boldsymbol{\Omega}_{t-k}$, however, it resembles the $\boldsymbol{\Sigma}_t$ process implied by a multivariate GARCH model. Consider the Engle and Kroner (1995) model as an example. It takes the form

$$\boldsymbol{e}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{z}_t \tag{9}$$

with

$$\boldsymbol{\Sigma}_t = \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{B}'\boldsymbol{\Sigma}_{t-1}\boldsymbol{B} + \boldsymbol{A}'\boldsymbol{e}_{t-1}\boldsymbol{e}'_{t-1}\boldsymbol{A}, \tag{10}$$

where $\boldsymbol{C}$ is a lower-triangular $N \times N$ parameter matrix, $\boldsymbol{B}$ and $\boldsymbol{A}$ are $N \times N$ parameter matrices, and $\boldsymbol{z}_t$ is an $N \times 1$ vector of uncorrelated standard normal random variables. If we let $\boldsymbol{C} = \boldsymbol{0}$, $\boldsymbol{B} = \mathrm{diag}\{b_1, \dots, b_N\}$, and $\boldsymbol{A} = \mathrm{diag}\{a_1, \dots, a_N\}$, then we can express Eq. (10) as

$$\boldsymbol{\Sigma}_t = \boldsymbol{B}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{B} \odot \boldsymbol{\Sigma}_{t-1} + \boldsymbol{A}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{A} \odot \boldsymbol{e}_{t-1}\boldsymbol{e}'_{t-1}, \tag{11}$$

where $\boldsymbol{1}$ denotes an $N \times 1$ vector of ones.

Comparing Eqs. (8) and (11) reveals that we can nest the rolling estimator under the GARCH model by setting $\boldsymbol{\Omega}_{t-k} = \boldsymbol{B}^{k-1}\boldsymbol{A}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{A}\boldsymbol{B}^{k-1}$. Similarly, we can treat the fitted values of $\boldsymbol{\Sigma}_t$ from the GARCH model as rolling estimates of this matrix for a specific choice of $\boldsymbol{\Omega}_{t-k}$. This relation motivates a GARCH-based benchmark for our empirical analysis, and it also suggests a simple approach for selecting an optimal weighting scheme for our rolling estimator.

3.1.1. Optimal weighting scheme

As in FKO (2001) our weighting scheme is of the form $\boldsymbol{\Omega}_{t-k} = \alpha \exp(-\alpha k)\boldsymbol{1}\boldsymbol{1}'$. Under this weighting scheme, Eq. (8) becomes

$$\hat{\boldsymbol{\Sigma}}_t = \exp(-\alpha)\hat{\boldsymbol{\Sigma}}_{t-1} + \alpha \exp(-\alpha)\boldsymbol{e}_{t-1}\boldsymbol{e}'_{t-1}. \tag{12}$$

This choice of weights reflects a number of considerations. First, it is consistent with Foster and Nelson (1996), who show that exponentially weighted estimators generally produce the smallest asymptotic mean squared error (MSE). Second, it guarantees that $\hat{\boldsymbol{\Sigma}}_t$ is positive definite, which is essential in our portfolio optimization application. Third, it implies that a single parameter (α) controls the rate at which the weights decay with the lag length. This parsimony facilitates our sensitivity analysis.

To estimate the optimal decay rate, we use a loss function motivated by the relation between rolling estimators and GARCH models. Our strategy is to recast the rolling estimator as a restricted multivariate GARCH model, treat α as an unknown parameter, and then estimate its value by maximum likelihood. Accordingly, we define the optimal decay rate to be the value of α that maximizes the likelihood function for the model

$$\boldsymbol{e}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{z}_t \tag{13}$$

with

$$\boldsymbol{\Sigma}_t = \exp(-\alpha)\boldsymbol{\Sigma}_{t-1} + \alpha \exp(-\alpha)\boldsymbol{e}_{t-1}\boldsymbol{e}'_{t-1}, \tag{14}$$

where $\boldsymbol{z}_t \sim \mathrm{NID}(\boldsymbol{0}, \boldsymbol{I})$. Using this approach facilitates a comparison between the empirical performance of our rolling estimator and the performance of a multivariate GARCH model. This allows us to determine whether our rolling estimator is a reasonable benchmark for a class of more complex models based on daily returns.

Of course, by estimating the optimal decay rate using the full dataset, we could be introducing a look-ahead bias into our results. We do not expect this to be a problem given the evidence in FKO (2001). They report that the optimal decay rate under a statistical loss function is very different from the decay rate that maximizes the empirical performance of the volatility-timing strategies. This is likely to be the case here as well. Moreover, in the empirical analysis, we explicitly examine how the choice of decay rate affects our results. We simply use the estimated decay rate to establish a baseline from which we begin this analysis.

3.2. Rolling estimators based on intradaily returns

The work of ABDL (2001) and Barndorff-Nielsen and Shephard (2001, 2002) suggests that we can use intradaily returns to construct volatility estimators that are more efficient than those based on daily returns. To illustrate the argument, suppose that log prices are generated by a multivariate continuous-time stochastic volatility process and let $\boldsymbol{S}_t$ denote the time t value of the associated $N \times N$ positive-definite diffusion matrix. In this case, it is natural to use the integrated diffusion matrix $\int_0^1 \boldsymbol{S}_{t+\tau}\, \mathrm{d}\tau$ as a measure of the latent covariance matrix of the vector of continuously compounded returns over the interval from t to $t+1$.

Now suppose we divide this interval into n subperiods of length h and let $\boldsymbol{r}_{t+jh}$ denote the vector of continuously compounded returns over the subperiod that ends at time $t+jh$. ABDL (2001) show that, under weak regularity conditions, the theory of quadratic variation implies that $\sum_{j=1}^{n} \boldsymbol{r}_{t+jh}\boldsymbol{r}'_{t+jh} - \int_0^1 \boldsymbol{S}_{t+\tau}\, \mathrm{d}\tau \to \boldsymbol{0}$ almost surely as $h \to 0$. It follows, therefore, that by using intradaily returns, we can construct nonparametric estimates of the integrated covariance matrix that are asymptotically consistent. We refer to the quantity $\boldsymbol{V}_{t+1} = \sum_{j=1}^{n} \boldsymbol{r}_{t+jh}\boldsymbol{r}'_{t+jh}$ as the realized covariance matrix for the interval t to $t+1$. This terminology reflects the fact that the integrated diffusion matrix effectively becomes observable as $h \to 0$.

To exploit these results, we need to encompass the realized volatility approach within our rolling estimator framework. A simple way to do this is to replace the outer product of the vector of return innovations with the realized covariance matrix when forming our rolling estimator. In other words, we replace Eq. (12) with

$$\tilde{\boldsymbol{\Sigma}}_t = \exp(-\alpha)\tilde{\boldsymbol{\Sigma}}_{t-1} + \alpha \exp(-\alpha)\boldsymbol{V}_{t-1}, \tag{15}$$

and estimate the optimal decay rate as before. Andreou and Ghysels (2002) study similar estimators in a univariate setting. Intuitively, $\tilde{\boldsymbol{\Sigma}}_t$ should be more efficient than $\hat{\boldsymbol{\Sigma}}_t$ because the realized variances and covariances provide more precise estimates of the integrated variances and covariances than the squares and cross-products of the daily return innovations. Andersen and Bollerslev (1998), for instance, show that the variance of daily return innovations can easily be an order of magnitude larger than the variance of cumulative squared intradaily returns.

3.2.1. Bias corrections

Although using intradaily returns leads to efficiency gains, it can also impart significant biases. In our case, these biases arise from three sources: a lack of intradaily return observations when markets are closed overnight; nonsimultaneous price observations across markets; and serial correlation induced by price discreteness and bid-ask bounce. We can mitigate the impact of the last two factors by choosing the sampling interval judiciously, so the lack of intradaily returns for the overnight period is our biggest concern. Not only does this cause the realized variances and covariances to be biased toward zero, it also results in a loss of information about the dynamics of the conditional covariance matrix of daily returns.

We address the information loss and the bias problems in two steps. To reduce the information loss, we include the overnight returns when constructing the realized covariance matrix. This is accomplished by including the outer product of the vector of overnight returns as another term in the summation that determines $\boldsymbol{V}_{t-1}$ in Eq. (15). Of course, this approach is only a partial solution. The outer product of the vector of overnight returns is an imprecise estimator of the integrated covariance matrix over the nontrading period. Nonetheless, the lack of precision is likely to be more than offset by the information gains.

To deal with the bias problem, we apply two simple corrections. We construct these corrections using contemporaneous estimates from the daily-returns-based rolling estimator because these should be largely unaffected by the biases. For the realized variances, a simple backward-looking measure of the bias at time t for the ith contract is

$$\beta_{it} = \frac{\sum_{l=1}^{q} \hat{\sigma}^2_{it-l}}{\sum_{l=1}^{q} \tilde{\sigma}^2_{it-l}}, \tag{16}$$

where $\hat{\sigma}^2_{it}$ and $\tilde{\sigma}^2_{it}$ denote the ith diagonal elements of $\hat{\boldsymbol{\Sigma}}_t$ and $\tilde{\boldsymbol{\Sigma}}_t$, respectively. We minimize the impact of the bias by using β_{it} to construct a bias-corrected version of

$\tilde{\sigma}_{it}^2$ for all i. Specifically, we replace the ith diagonal element of $\tilde{\boldsymbol{\Sigma}}_t$ with

$$\tilde{\sigma}_{it}^{*2} = \beta_{it}\tilde{\sigma}_{it}^2, \tag{17}$$

to obtain the bias-corrected variance estimates.

Next, consider the realized covariances. To correct for the bias, we replace the off-diagonal elements of $\tilde{\boldsymbol{\Sigma}}_t$ with

$$\tilde{\sigma}_{ijt}^{*} = \tilde{\sigma}_{it}^{*}\tilde{\sigma}_{jt}^{*}(\tilde{\rho}_{ijt} + \beta_{ijt}), \tag{18}$$

where

$$\beta_{ijt} = \frac{1}{q}\sum_{l=1}^{q}(\hat{\rho}_{ijt-l} - \tilde{\rho}_{ijt-l}), \tag{19}$$

$\hat{\rho}_{ijt} = \hat{\sigma}_{ijt}/(\hat{\sigma}_{it}\hat{\sigma}_{jt})$, and $\tilde{\rho}_{ijt} = \tilde{\sigma}_{ijt}/(\tilde{\sigma}_{it}\tilde{\sigma}_{jt})$. Here we use an additive correction because, unlike in Eq. (16), the moving averages used to construct our bias measure can be positive, negative, or zero. We also impose the constraint $-1 \leqslant \tilde{\rho}_{ijt} + \beta_{ijt} \leqslant 1$.

In Eqs. (16) and (19), the choice of q reflects a tradeoff between two opposing considerations. If the biases are constant, then we should make q as large as possible to maximize the precision of our bias estimates. If, as seems more likely, the biases vary with the level of trading activity and/or market volatility, then a small value of q may be more effective. Fortunately, the precise choice of q does not appear to be critical as long as it is small enough to capture the time variation in the biases. We experimented with values from five to 504 days with little apparent effect on our empirical results. All of the results reported below are based on $q = 22$ days.

4. Data and preliminary empirical analysis

Our empirical analysis is based on the same futures contracts studied by FKO (2001): S&P 500 futures (Chicago Mercantile Exchange), Treasury bond futures (Chicago Board of Trade), and gold futures (New York Mercantile Exchange). The sample period is January 3, 1984 to November 30, 2000. We obtain daily closing prices for gold futures from Datastream International and intraday transactions data for each contract from the Futures Industry Institute. The S&P 500 futures data prior to 1994 are obtained directly from the CME. We exclude from the dataset all days on which any of the three futures markets is closed.

To construct the return series, we generally use the nearby contract in each market. However, we switch to the second nearby contract for stocks when the nearby contract is in its final week and for bonds and gold when the nearby contract is in the delivery month. The gold futures contract closes at 1:30 pm whereas the bond and stock contracts close at 2:00 pm and 3:15 pm (all central standard times). Therefore, we assume that our portfolios are rebalanced at 1:30 pm each day, and that these trades occur at the closing price for gold and at the last transaction prices before 1:30 pm for stocks and bonds. We refer to these as the "1:30 prices". This procedure is the same as in FKO (2001).

4.1. Conditional covariance matrix estimates based on daily returns

To implement our rolling estimator based on daily returns, we use returns computed from the 1:30 prices. Summary statistics for these returns are shown in Panel A of Table 1. As explained earlier, we estimate the optimal decay rate by fitting the GARCH analog of the rolling estimator to the daily returns. This yields an

Table 1
Summary statistics for daily and five-minutes returns
The table provides summary statistics for the daily and five-minutes returns for stock, bond, and gold futures. The daily returns are based on closing prices for gold and 1:30 prices for stocks and bonds for days that all three markets are open. The five-minutes returns are based on intraday transaction prices nearest to each five-minutes gridpoint. We exclude gridpoints at the beginning and end of each day until the nearest transaction is less than two and a half minutes away. For the five-minutes daily returns, we report the number of days, the unconditional mean returns and volatilities (both annualized), and the unconditional correlations. For the five-minutes returns (r_t), we report the number of days and statistics for the number of transaction prices per day, the number of 5-min returns per day, and the autocorrelations of r_t and $|r_t|$. The sample period is January 3, 1984 to November 30, 2000. The five-minutes returns are continuously compounded and exclude October 19–30, 1987.

Statistic		Stocks	Bonds	Gold
Panel A: daily returns				
Days		4,238	4,238	4,238
Mean return		0.099	0.065	−0.066
Volatility		0.169	0.101	0.143
Correlations	Stocks		0.326	−0.124
	Bonds			−0.168
Panel B: five-minutes returns				
Days		4,228	4,228	4,119
Trades per day	Minimum	241	210	42
	Average	2,686	1,594	1,101
	Maximum	5,228	3,719	4,646
Returns per day	Minimum	20	24	12
	Average	80	77	72
	Maximum	87	87	78
Autocorrelation of r_t	Lag 1	−0.010	−0.077	−0.053
	Lag 2	−0.021	0.011	−0.023
	Lag 3	−0.007	0.001	−0.004
	Lag 4	−0.002	0.004	−0.001
	Lag 5	0.000	0.002	−0.004
Autocorrelation of $\|r_t\|$	Lag 1	0.289	0.223	0.314
	Lag 2	0.264	0.174	0.228
	Lag 3	0.247	0.128	0.187
	Lag 4	0.232	0.095	0.155
	Lag 5	0.223	0.072	0.131

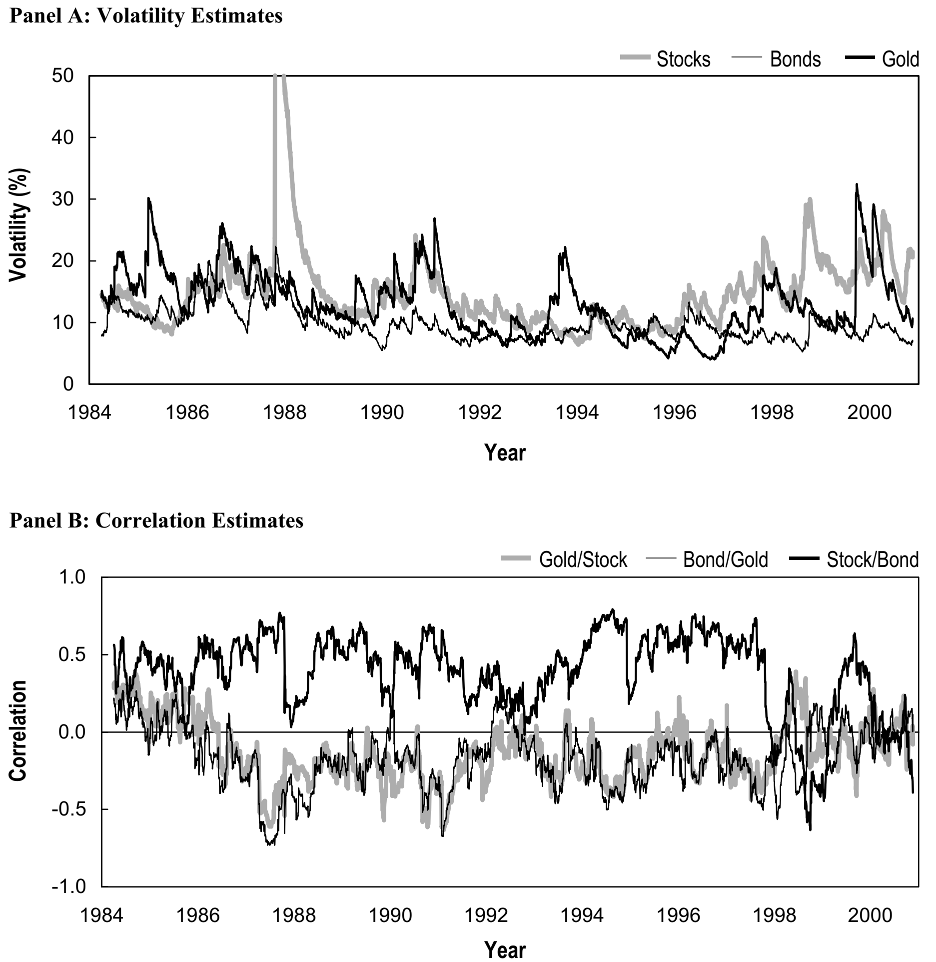

Fig. 1. Daily conditional volatility and correlation estimates for stock, bond, and gold futures using the daily-returns-based rolling estimator. The estimates are generated using a decay rate of $\alpha = 0.031$. Panel A shows the volatility estimates for each market (reported as annualized percentages) and Panel B shows the implied correlations for each pair of markets based on the conditional variance and covariance estimates. The sample period is January 3, 1984 to November 30, 2000.

estimate of $\alpha = 0.031$ with a log likelihood value of $-6,494.5$. Fig. 1 plots the conditional covariance matrix estimates using this decay rate. Panel A shows the estimates of the daily return volatility for each contract and Panel B shows the estimates of the correlation between the daily returns for each pair of contracts.

4.2. Conditional covariance matrix estimates based on intradaily returns

We now implement our rolling estimator based on realized volatility. We construct the intradaily returns used to implement this estimator using a five-minute sampling frequency. This sampling frequency balances the benefits derived from large sample asymptotics and the detrimental impact of market microstructure effects (ABDE, 2001; ABDL, 2001; Areal and Taylor, 2002). To account for serial correlation induced by microstructure effects, ABDE (2001) suggest prewhitening the five-minute returns by fitting a moving average model. With this approach, however, the resulting conditional covariance matrix estimates would depend on the entire dataset. Since we want to focus as much as possible on out-of-sample results, we use an interpolation procedure developed by ABDL (2001).

4.2.1. Computing the five-minute returns

For each day in the sample, we compute the series of intradaily returns for each contract using the following procedure. First, we construct a grid of five-minute intervals that spans the trading day. Next, we identify the transaction prices that straddle each of the grid points, deleting any grid points that are more than two and a half minutes outside the first and last transactions of the day. Finally, we use linear interpolation to estimate the log price at each grid point and then we take first differences of these prices to obtain the continuously compounded five-minute returns.

We do not construct intradaily returns for the period immediately surrounding the 1987 stock market crash. This decision is motivated by a number of factors: gold and bonds are missing data for a couple days, price limits went into effect for bonds, the minimum tick size for stocks increased from 0.05 to at least 1.00 for several days, and the stock market often closed early. Since an investor trying to compute realized variances and covariances for one of these days would likely conclude that the intraday data were unreliable, we substitute daily returns to construct these measures from October 19–30, 1987.

Panel B of Table 1 provides summary statistics for the five-minute returns. The average number of returns per day ranges from 72 for gold to 80 for stocks. The return autocorrelations for stocks are close to zero, but they are negative and significant at lag one for both bonds and gold. This is likely due to market microstructure effects. While all of the markets are active (each averages over 1,000 trades per day), the bond and gold markets are noticeably less active than the stock market. Therefore, it should not be surprising that our interpolation procedure fails to completely purge the negative serial correlation in the bond and gold returns. The autocorrelations for the absolute returns are all positive and significant. This shows the level of volatility persistence in the five-minute returns.

4.2.2. Computing the realized variances and covariances

To compute the realized covariance matrix for the interval $t-1$ to t, we use the five-minute returns from 1:30 on day $t-1$ to 1:30 on day t. We construct the realized variances using all of these returns. For the realized covariances, however, we can use only those returns that are contemporaneous across markets. For example, the returns after 1:30 pm in the stock and bond markets cannot be used to compute the realized covariances with gold because the gold market is closed. Since this reduces the number of available observations, we construct the realized covariances using the following two-step procedure. First, we use the contemporaneous returns to compute a preliminary set of realized volatilities and covariances from which we infer the realized correlations. Second, we convert these correlations back into covariances using the realized volatilities based on the entire set of five-minute returns.

Panel A of Figs. 2 and 3 plot the realized volatilities and correlations. In general, they appear more variable than the realized volatilities and correlations in ABDL (2001), but this should not be surprising. ABDL (2001) use quote mid-points from the currency markets where trading occurs around the clock, giving them 288 five-minute returns. They also exclude certain holidays and other days where intradaily data are incomplete. In contrast, we use transaction prices, have a maximum of 87 five-minute returns, and we include days with incomplete data.

4.2.3. The rolling estimates based on realized volatility

For the rolling estimator based on realized volatility, our GARCH approach for obtaining the optimal decay rate yields an estimate of $\alpha = 0.064$ and a log likelihood value of $-6,428.8$.[4] This decay rate is larger than that obtained using daily returns and the log likelihood value is higher as well. Both of these results are consistent with prior evidence regarding the realized volatility approach. A larger decay rate corresponds to larger weights on more recent data and the higher log likelihood suggests a gain in efficiency.

Panel B of Figs. 2 and 3 plot the rolling estimates based on the realized variances and covariances using a decay rate of 0.064. As expected, these estimates appear biased toward zero in comparison to the rolling estimates based on daily returns (see Fig. 1). We correct for this bias in two steps as described in Section 3.2.1. First, we include overnight returns when constructing the realized variances and covariances. The resulting rolling estimates, plotted in Panel C of Figs. 2 and 3, indicate that including overnight returns reduces the bias but not to a substantial degree. We obtain our final set of estimates by applying our variance and covariance bias corrections. Panel D of Figs. 2 and 3 plot these estimates which look very similar to the rolling estimates based on daily returns.

5. Empirical results

In this section, we evaluate the performance of the volatility-timing strategies and assess the incremental value of using the realized volatility approach to construct the conditional covariance matrix estimates. To implement the volatility-timing strategies, we need to supply an estimate of the vector of unconditional expected returns. FKO (2001) argue that using a single estimate of this vector is problematic because estimation risk is likely to be a significant concern. Instead, they consider a range of inputs which are generated via a bootstrap procedure (Efron, 1979). We use a similar approach here. We start with the limiting case where there is no estimation risk, i.e., the vector of ex post mean returns is known. Then we introduce various

[4] These estimates are based on our adjusted realized variances and covariances which include overnight returns and the bias corrections described in Section 3.2.1.

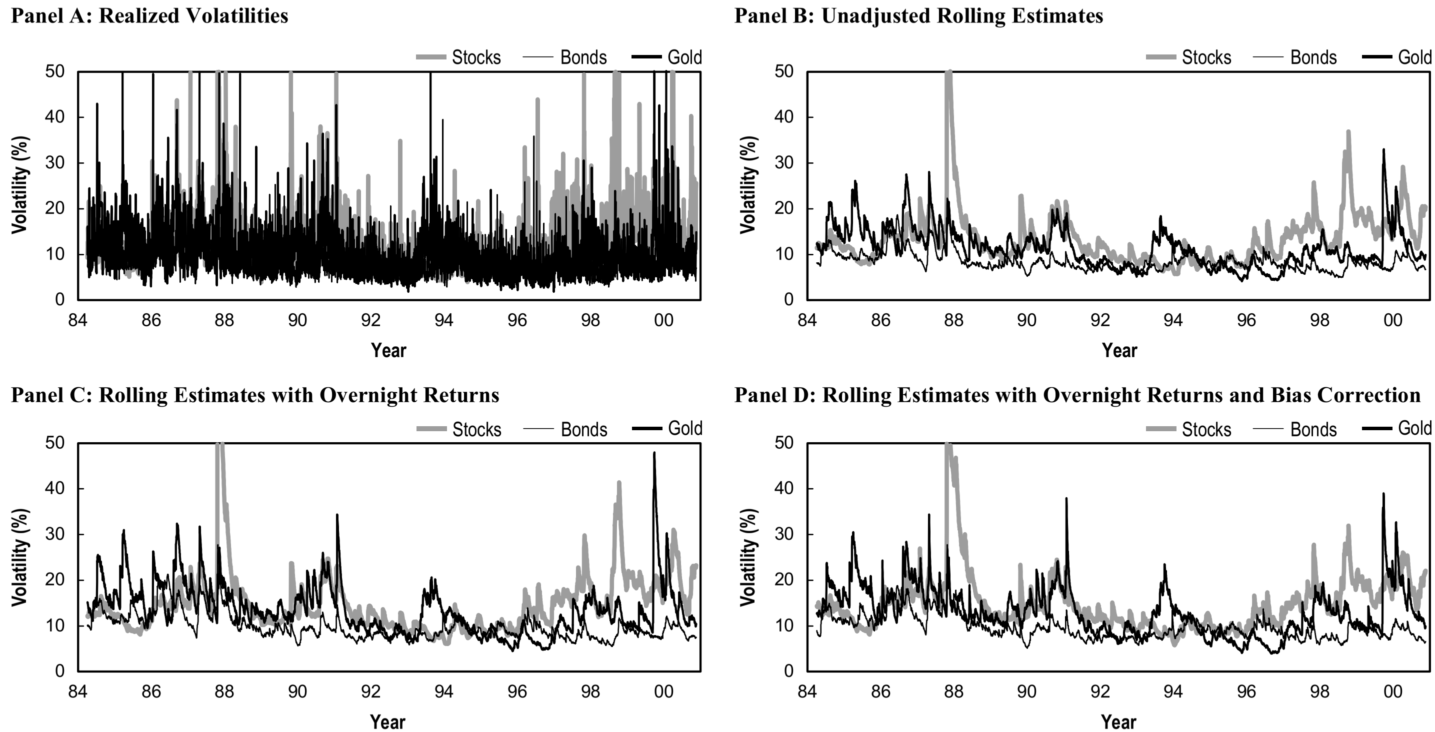

Fig. 2. Daily conditional volatility estimates for stock, bond, and gold futures using the realized volatility approach. Panel A shows the cumulative squared five-minute returns for each day in the sample. Panel B shows the unadjusted rolling estimates with a decay rate of $\alpha = 0.064$. Panel C shows the rolling estimates after including the overnight (i.e., close-to-open) returns. Panel D shows the rolling estimates after adding the volatility bias correction. All of the volatility estimates are shown as annualized percentages. The sample period is January 3, 1984 to November 30, 2000.

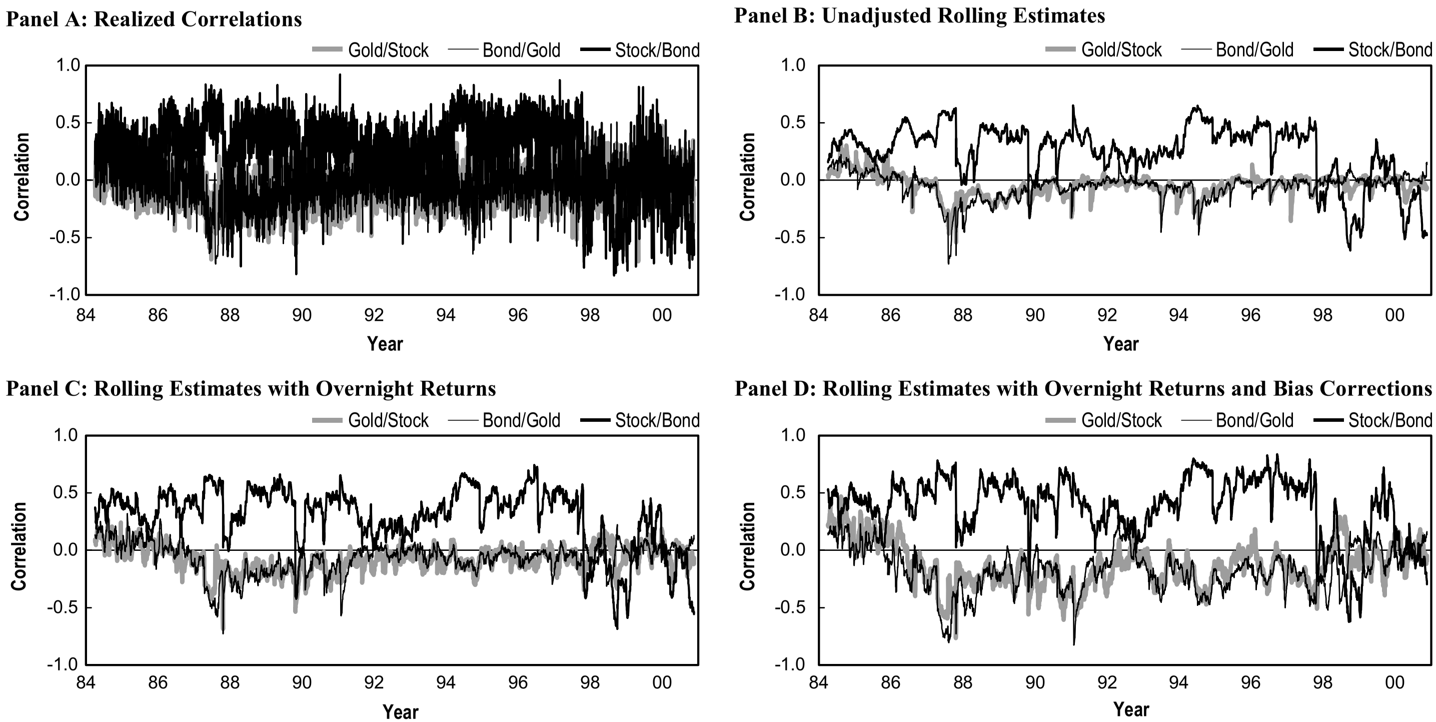

Fig. 3. Daily conditional correlation estimates for stock, bond, and gold futures using the realized volatility approach. All of the correlation estimates are implied based on variance and covariance estimates. Panel A shows realized correlations based on cumulative squared five-minute returns and return cross-products. Panel B shows the unadjusted rolling estimates with a decay rate of $\alpha = 0.064$. Panel C shows the rolling estimates after including the overnight (i.e., close-to-open) returns. Panel D shows the rolling estimates after adding the volatility and correlation bias corrections. The sample period is January 3, 1984 to November 30, 2000.

levels of estimation risk by using a version of the block bootstrap advocated by Hall and Horowitz (1996). Finally, we provide some insights into the source of the performance gains associated with switching to the realized volatility approach.

5.1. The case of no estimation risk

Fig. 4 plots the optimal portfolio weights using the rolling covariance matrix estimator based on realized volatility. The weights for the minimum volatility strategy (Panel A) assume a target expected return of 10% and the weights for the maximum return strategy (Panel B) assume a target volatility of 12%. Although

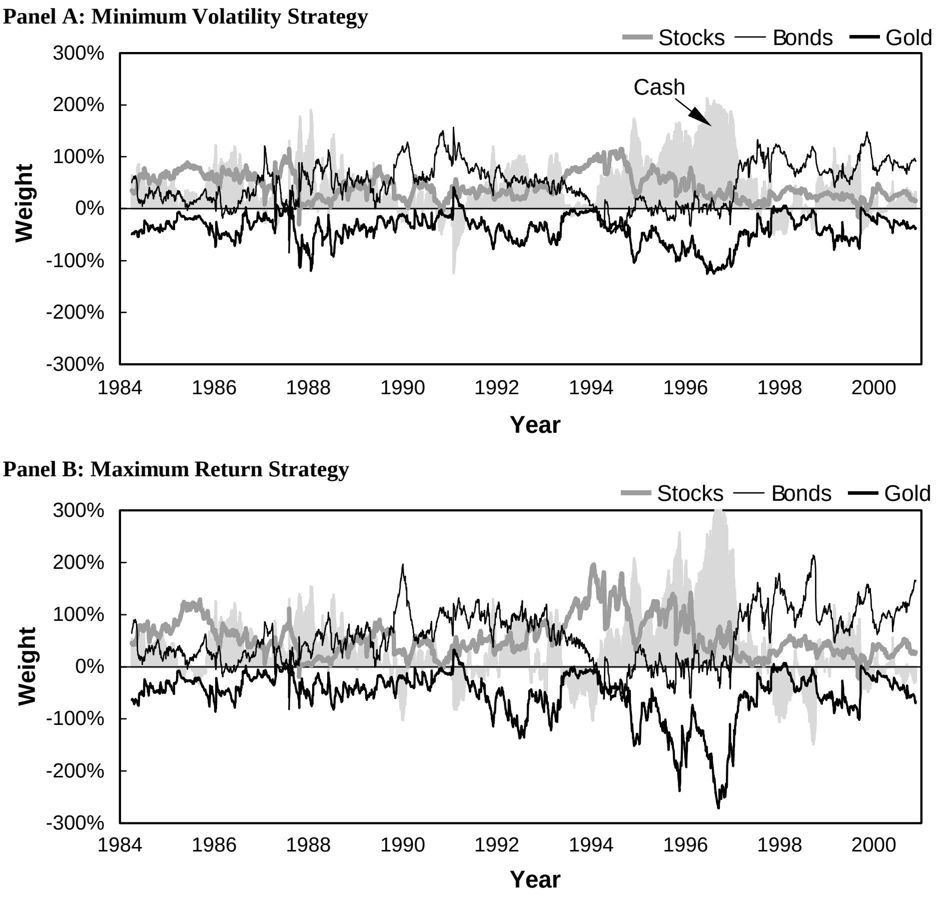

Fig. 4. Portfolio weights from daily mean-variance optimizations using stock, bond and gold futures for the case of no estimation risk. The optimizations assume that the expected return for each asset is constant and equal to its in-sample mean and that the conditional covariance matrix equals our daily out-of-sample estimate based on the realized volatility approach. Panel A shows the portfolio weights that minimize conditional volatility while setting the expected return equal to 10%, and Panel B shows the weights that maximize expected return while setting the conditional volatility equal to 12%.

these targets are arbitrary, changing them simply alters the allocation between the risky assets and cash and has little effect on our results. The portfolio weights for both strategies tend to be positive for stocks and bonds and negative for gold, consistent with the mean returns for these assets over our sample. Also, as expected, the variation in the weights over time tracks the variation in the covariance matrix estimates shown in Figs. 2 and 3. The weights are generally more extreme for the maximum return strategy because it tends to plot further to the right on the efficient frontier. The optimal portfolio weights using the rolling covariance matrix estimator based on daily returns (not shown) are similar but they exhibit greater variability because the covariance matrix estimates (Fig. 1) are less precise.

Table 2 compares the empirical performance of the volatility-timing strategies using the two different rolling estimators of the conditional covariance matrix. For the minimum volatility (maximum return) strategy, using the rolling estimator based on daily returns yields an average return of 10.5% (12.7%), a sample volatility of 11.4% (13.0%), and a Sharpe ratio of 0.92 (0.98). Switching to the rolling estimator based on realized volatility, the average return increases to 11.7% (14.2%), the volatility is comparable at 11.3% (13.0%), and the Sharpe ratio increases to 1.03 (1.09). The table also reports results for each three-year subperiod and for two subperiods that exclude the 1987 crash. These results indicate that

Table 2
Performance of the volatility-timing strategies with no estimation risk
The table summarizes the performance of the volatility-timing strategies for the case of no estimation risk regarding expected returns. Each day, we solve a portfolio optimization problem in which the expected return for each asset equals its in-sample mean return and the conditional covariance matrix is estimated out of sample using either the daily-returns-based rolling estimator ("daily") or the realized-volatility-based rolling estimator ("realized"). In each case, we solve for the portfolio weights that minimize conditional volatility subject to a target expected return of 10% ("minimum volatility strategy") and the weights that maximize expected return subject to a target volatility of 12% ("maximum return strategy"). For each set of weights, we report the mean next-day return (μ), volatility (σ), and Sharpe ratio (SR). The means and volatilities are expressed as annualized percentages. The sample period is January 3, 1984 to November 30, 2000. We also report the results for each three-year subperiod and for two non-crash periods that exclude either October 19–30, 1987 or the entire 1986–1988 subperiod.

		Minimum volatility strategies						Maximum return strategies					
		Daily			Realized			Daily			Realized		
Period	Obs.	μ	σ	SR	μ	σ	SR	μ	σ	SR	μ	σ	SR
Entire sample	4,175	10.5	11.4	0.92	11.7	11.3	1.03	12.7	13.0	0.98	14.2	13.0	1.09
1984–1985	440	16.3	10.0	1.63	16.4	10.0	1.64	20.4	12.9	1.58	21.2	13.1	1.62
1986–1988	757	9.7	16.6	0.59	14.0	16.4	0.85	6.0	13.5	0.45	9.3	13.5	0.69
1989–1991	758	10.7	11.8	0.91	11.0	11.8	0.93	10.4	13.1	0.79	11.7	12.9	0.91
1992–1994	751	5.0	8.3	0.60	5.6	8.3	0.68	7.6	12.7	0.59	8.9	12.7	0.70
1995–1997	744	18.4	8.9	2.07	19.3	8.9	2.17	28.0	12.9	2.17	29.3	12.9	2.26
1998–2001	726	5.3	10.1	0.52	5.8	10.1	0.58	7.4	12.9	0.57	7.8	12.9	0.60
Ex Oct 19–30, 1987	4,165	10.0	11.1	0.90	10.8	11.1	0.97	12.5	12.9	0.97	13.6	12.9	1.05
Ex 1986–1988	3,418	10.7	9.9	1.08	11.2	9.9	1.13	14.2	12.9	1.10	15.2	12.9	1.18

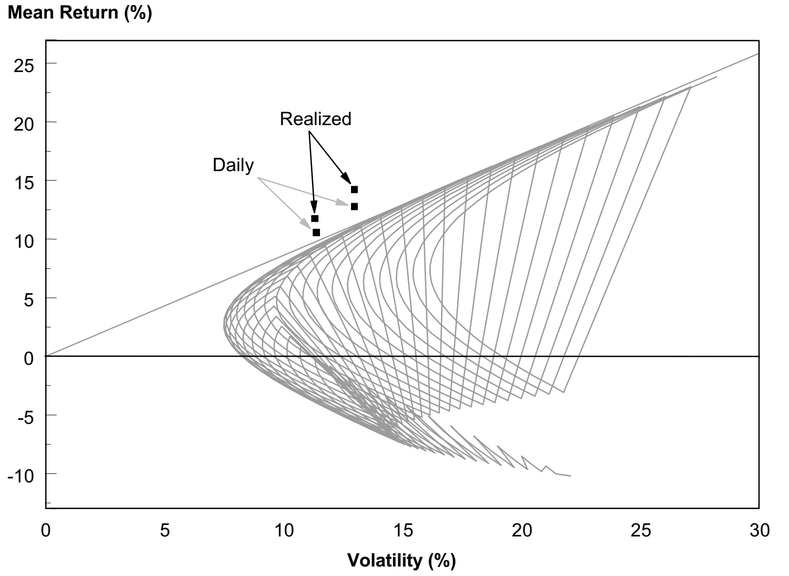

Fig. 5. The performance of the volatility-timing strategies versus the ex post efficient frontier using static portfolio weights. We plot the return-volatility combinations earned by the volatility-timing strategies using rolling conditional covariance matrix estimates based on either daily returns ("daily") or realized volatility ("realized"). We also plot the set of return-volatility combinations that could be earned during our sample period using static portfolio weights in stock, bond, and gold futures. The volatility-timing strategies are based on the ex post means and ex ante estimates of the conditional covariance matrix while the static portfolios are based on ex post means and ex post variances and covariances. The returns and volatilities are annualized on the basis of 252 trading days per year.

switching to the realized-volatility-based estimator generates performance gains in each subperiod.

Fig. 5 provides evidence on the value of volatility timing in the case of no estimation risk. The figure shows the ex post minimum-variance frontier for static portfolios over our sample period. The highest Sharpe ratio achieved by one of these portfolios is 0.81. In contrast, the dynamic volatility-timing strategies earn Sharpe ratios ranging from 0.92 to 1.09. Thus, the volatility-timing strategies, which are based on the ex post means and ex ante estimates of the conditional covariance matrix, dominate the unconditionally efficient static portfolios which are based on both ex post means and ex post variances and covariances. This suggests that volatility timing can generate economic value. However, to assess whether this value is realizable we need to incorporate the effects of estimation risk.

5.2. Controlling for estimation risk

In practice, an investor implementing the volatility-timing strategies faces uncertainty about expected returns. We use a block bootstrap approach to control for this uncertainty and assess its impact on the results. Suppose, for example, that we want to mimic the estimation risk inherent in using a sample size comparable to ours to estimate the expected returns. First, we form an artificial sample of $k = 4,000$ returns by randomly sampling, with replacement, nonoverlapping blocks of 15 observations from the series of actual returns.[5] Next, we use the mean returns from this artificial sample, together with our conditional covariance matrix estimates, to compute the portfolio weights for the volatility-timing strategies. Finally, we apply these weights to the actual returns and evaluate the performance of the strategies. Repeating this procedure over 1,000 trials, we can assess how the volatility-timing strategies perform across a wide range of plausible inputs for the vector of expected returns.

We measure the value of volatility timing by comparing the performance of the volatility-timing strategies to that of an ex ante unconditionally efficient static portfolio. To construct this portfolio, we use the artificial sample in each bootstrap trial to estimate the unconditional mean and unconditional covariance matrix of daily returns. Then, using these estimates, we solve for the fixed portfolio weights that either minimize volatility or maximize expected return subject to the same target return or volatility as the volatility-timing strategy. Since the static portfolio is based on the same level of estimation risk as the volatility-timing strategy, their relative performance reflects the gains attributable to volatility timing.

Although our bootstrap experiment mimics the uncertainty faced by an investor in practice, a direct application of the procedure ignores potentially valuable non-sample information about the parameters. Specifically, asset pricing theory suggests that $\mu_s > \mu_b > \mu_g$, $\mu_s > 0$, $\mu_b > 0$, and $\sigma_s > \sigma_b$, where μ and σ denote the unconditional expected return and unconditional volatility and the subscripts denote stocks (s), bonds (b), and gold (g). By incorporating these inequalities in the estimation procedure, an investor can potentially reduce estimation risk. We approximate this strategy by treating the inequalities as exclusionary criteria: if the sample means or volatilities for a particular trial violate any of the criteria, we discard the estimates and generate a new bootstrap sample. We repeat this process until we have a total of 1,000 trials in which the inequalities are satisfied.[6]

Table 3 shows the distributional properties of the mean and covariance matrix estimates obtained from our bootstrap procedure. The table reports the average of the estimates across trials and the associated standard errors. For example, using an artificial sample size of $k = 1,000$, the average expected return estimates for stocks, bonds, and gold are 14.4%, 7.1%, and −7.6% annualized. As k increases, the expected return estimates converge to the actual sample means for our dataset. However, even with $k = 10,000$, the standard errors are fairly large. Thus, it is clear that using our bootstrap approach generates a wide range of estimates for the expected returns.

As expected, some biases arise as a result of applying our exclusionary criteria in the bootstrap trials. Using $k = 1,000$, for example, 45% of the total samples

[5] The choice of block length is not critical. We experimented with lengths from 1 to 25 days and found little change in the average parameter estimates or the associated standard errors.

[6] This approach is similar in spirit to Bayesian techniques for exploiting non-sample information. If a Bayesian investor incorporates inequality constraints in his prior, then the posterior distribution assigns zero probability to regions of the parameter space in which the constraints are violated.

Table 3

Estimates of the unconditional means, volatilities, and correlations in the bootstrap trials

The table summarizes the distributional properties of the unconditional mean return and covariance matrix estimates obtained from our bootstrap trials. In each trial, we generate an artificial sample of k returns by randomly sampling with replacement from the actual returns, selecting nonoverlapping blocks of 15 observations. We then use the artificial sample to estimate the unconditional mean returns and the unconditional covariance matrix. We discard trials that violate any of our exclusionary criteria: $\mu_s > \mu_b > \mu_g$, $\mu_s > 0$, $\mu_b > 0$, and $\sigma_s > \sigma_b$, where μ and σ are the mean return and volatility estimates for stocks (s), bonds (b), and gold (g). The results reported in the table are the average estimates based on 1,000 acceptable trials. We report the average mean return and volatility estimates for each asset (as annualized percentages) and the average correlation estimates (ρ) for each pair of assets. Standard errors are reported below the averages.

		Mean returns			Covariance matrix					
k	Trials excluded	μ_s	μ_b	μ_g	σ_s	σ_b	σ_g	ρ_{sb}	ρ_{sg}	ρ_{bg}
1,000	45.3%	14.40	7.10	−7.63	15.93	10.06	14.09	0.346	−0.126	−0.166
		5.36	3.80	6.07	1.97	0.50	1.05	0.061	0.046	0.049
2,000	33.6%	12.56	6.46	−6.98	16.21	10.08	14.17	0.340	−0.125	−0.168
		4.23	3.04	4.65	1.63	0.36	0.77	0.045	0.032	0.036
3,000	26.5%	11.62	6.16	−6.87	16.41	10.10	14.20	0.336	−0.125	−0.167
		3.67	2.65	3.91	1.43	0.30	0.65	0.036	0.027	0.030
4,000	21.7%	11.13	6.07	−6.61	16.48	10.10	14.21	0.335	−0.124	−0.168
		3.30	2.32	3.40	1.24	0.26	0.54	0.032	0.022	0.026
5,000	17.4%	10.91	6.13	−6.80	16.53	10.11	14.23	0.333	−0.125	−0.168
		2.85	2.22	3.03	1.13	0.24	0.50	0.028	0.020	0.023
10,000	9.9%	10.49	6.20	−6.51	16.68	10.12	14.23	0.330	−0.125	−0.168
		2.18	1.60	2.11	0.85	0.17	0.33	0.020	0.014	0.016

generated violate the criteria. Since the remaining 1,000 samples tend to be those with higher than average returns for stocks and bonds and lower than average returns for gold, the expected return estimates for stocks and bonds exhibit an upward bias and those for gold exhibit a downward bias. In addition, the volatility estimates exhibit a downward bias because the excluded samples tend to be those containing more extreme observations. Increasing k reduces these biases; however, they are still apparent using $k = 10,000$, with nearly 10% of the samples violating the exclusionary criteria. We consider later the impact of these biases on our results.

Table 4 evaluates the performance of the volatility-timing strategies for various levels of estimation risk. For the strategies using the daily-returns-based rolling estimator, volatility timing clearly outperforms the ex ante efficient static porfolio. For example, with $k = 10,000$, the average returns for the minimum volatility and static portfolios are comparable, but volatility timing generates lower volatility (11.5% versus 12.2%) and a higher Sharpe ratio (0.89 versus 0.82). The p-value of 1.00 indicates that the volatility-timing strategy yields a higher Sharpe ratio in 100% of the bootstrap trials. Turning to the performance fees (Δ_γ), an investor with low relative risk aversion ($\gamma = 1$) would be willing to pay 18 basis

Table 4
Comparison of the volatility-timing strategies and the ex ante efficient static portfolios
The table compares the performance of the volatility-timing strategies to that of the ex ante efficient static portfolios. The results for each line in the table are based on 1,000 simulation trials using artificial samples of k returns to estimate the unconditional expected returns and the unconditional covariance matrix. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios ("static") and by the volatility-timing strategies using conditional covariance matrix estimates generated by either the daily-returns-based rolling estimator ("daily") or the realized-volatility-based estimator ("realized"). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which each volatility-timing strategy earns a higher Sharpe ratio than the corresponding static portfolio, and the average annualized basis point fees (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the static portfolio to the volatility-timing strategy.

	Static			Daily						Realized					
k	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies															
1,000	7.2	9.8	0.73	7.2	9.1	0.79	0.87	8.4	75.8	7.8	9.1	0.86	0.92	70.2	139.8
2,000	8.4	11.1	0.76	8.4	10.4	0.82	0.93	12.9	95.4	9.1	10.4	0.89	0.96	86.6	171.6
3,000	9.0	11.7	0.77	9.1	11.0	0.83	0.95	16.3	102.6	9.9	11.0	0.91	0.97	99.7	188.6
4,000	9.5	12.1	0.79	9.5	11.3	0.85	0.96	15.9	103.4	10.4	11.3	0.93	0.98	105.3	196.3
5,000	9.5	12.0	0.80	9.6	11.2	0.86	0.98	17.1	101.2	10.5	11.2	0.94	0.99	109.2	197.4
10,000	10.1	12.2	0.82	10.2	11.5	0.89	1.00	18.0	100.7	11.2	11.4	0.98	1.00	121.0	210.4
Panel B: maximum return strategies															
1,000	9.3	12.7	0.73	11.0	13.1	0.85	0.94	169.7	134.6	12.2	13.0	0.94	0.95	286.6	254.3
2,000	9.5	12.5	0.76	11.4	13.1	0.87	0.96	187.3	127.9	12.6	13.0	0.97	0.98	308.1	251.4
3,000	9.6	12.4	0.77	11.7	13.1	0.89	0.98	200.0	124.1	12.9	13.0	0.99	0.99	326.0	252.7
4,000	9.7	12.3	0.79	11.8	13.0	0.91	0.99	205.0	121.2	13.1	13.0	1.01	0.99	333.2	251.8
5,000	9.8	12.3	0.80	12.0	13.0	0.92	0.99	209.9	121.7	13.3	13.0	1.02	0.99	341.5	255.5
10,000	10.0	12.2	0.82	12.3	13.0	0.95	1.00	219.1	120.7	13.7	13.0	1.05	1.00	354.8	258.5

points per year to switch from the static to the volatility-timing portfolio, while an investor with high relative risk aversion ($\gamma = 10$) would be willing to pay 101 basis points. The maximum return strategy also outperforms the static portfolio but in this case the gains come mostly from a large increase in the average returns (12.3% vs. 10.0%).

Reducing the sample size k increases the level of estimation risk. Naturally, this tends to reduce the effectiveness of the volatility-timing strategies. However, it also reduces the performance of the ex ante efficient static portfolios, so the gains to volatility timing are not that sensitive to the level of estimation risk. Even with $k = 1,000$, the minimum volatility (maximum return) strategy has a higher Sharpe ratio than the static portfolio in 87% (94%) of the trials, and the performance fees are not very different from those with $k = 10,000$.

Now consider the results for the volatility-timing strategies using the rolling estimator based on realized volatility. The value of volatility timing increases substantially. The minimum volatility strategy generates both a higher return and lower volatility than the ex ante optimal static portfolio, independent of the level of estimation risk. The Sharpe ratios range from 0.86 to 0.98, the p-values are all greater than 92%, and the performance fees are on the order of 100 to 200 basis points. For the maximum return strategy, the Sharpe ratios range from 0.94 to 1.05, the p-values exceed 95%, and the performance fees are 250 to 350 basis points. These results suggest that volatility timing using the covariance matrix estimates based on realized volatility generates substantial economic value.

Table 5 evaluates the incremental contribution of the realized volatility approach on the performance of the volatility-timing strategies. The means, volatilities, and Sharpe ratios are the same as those reported in Table 4, but here the p-values indicate the fraction of trials in which the rolling estimator based on realized volatility generates a higher Sharpe ratio than the estimator based on daily returns. These p-values are always nearly one. For the minimum volatility strategy, using realized volatility increases the average returns without much effect on volatility, so the performance fees are roughly the magnitude of this increase (62 to 110 basis points). For the maximum return strategy, using realized volatility generates both higher returns and lower volatility, and the performance fees are even greater (117 to 138 basis points). These results suggest that the incremental value of using realized volatility is substantial.

Table 5
Performance of the realized-volatility-based estimator vs. the daily-returns-based estimator
The table compares the performance of the volatility-timing strategies using a rolling estimator of the conditional covariance matrix based either on daily returns or on realized volatility. The results for each line in the table are based on 1,000 simulation trials using an artificial sample of k returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the volatility-timing strategies using either the daily-returns-based estimator ("daily") or the realized-volatility-based estimator ("realized"). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using realized volatility earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from using daily returns to using realized volatility.

	Daily			Realized					
k	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies									
1,000	7.2	9.1	0.79	7.8	9.1	0.86	0.99	61.8	64.0
2,000	8.4	10.4	0.82	9.1	10.4	0.89	1.00	73.7	76.3
3,000	9.1	11.0	0.83	9.9	11.0	0.91	1.00	83.4	86.0
4,000	9.5	11.3	0.85	10.4	11.3	0.93	1.00	89.4	92.9
5,000	9.6	11.2	0.86	10.5	11.2	0.94	1.00	92.1	96.3
10,000	10.2	11.5	0.89	11.2	11.4	0.98	1.00	103.0	109.7
Panel B: maximum return strategies									
1,000	11.0	13.1	0.85	12.2	13.0	0.94	0.99	116.9	119.7
2,000	11.4	13.1	0.87	12.6	13.0	0.97	1.00	120.8	123.5
3,000	11.7	13.1	0.89	12.9	13.0	0.99	1.00	126.0	128.6
4,000	11.8	13.0	0.91	13.1	13.0	1.01	1.00	128.2	130.6
5,000	12.0	13.0	0.92	13.3	13.0	1.02	1.00	131.6	133.7
10,000	12.3	13.0	0.95	13.7	13.0	1.05	1.00	135.7	137.7

Fig. 6 plots the distribution of performance fees across the bootstrap trials for $k = 4,000$. Panels A and B show the fees for the realized-volatility-based estimator relative to the ex ante efficient static portfolio. The distributions vary with the level of risk aversion but in every case almost all the mass is in the positive region. Panels C and D show the fees for switching from the rolling estimator based on daily returns to the one based on realized volatility. Again, almost all the mass of each distribution lies above zero, and the distributions are much tighter than those in Panels A and B. The implication is that using the more precise conditional covariance matrix estimates almost always leads to better volatility-timing performance.

Earlier we noted the biases that arise from applying the exclusionary criteria in our bootstrap trials. To assess their impact on our results, Table 6 examines the performance of the various strategies when we implement the bootstrap without any constraints. As expected, the performance of all the strategies deteriorates, especially with high levels of estimation risk, and the p-values are lower because the trials generate a wider range of parameter estimates than before. Nonetheless, the relative performance of the strategies remains unchanged. Switching to the realized-volatility-based estimator from either the unconditional or the daily-returns-based estimators generates roughly the same increase in Sharpe ratio and same performance fees as reported in Table 4. Moreover, if we directly compare the performance of the realized-volatility-based and daily-returns-based estimators (not shown), the realized-volatility-based estimator yields a higher Sharpe ratio in over 90% of the trials, regardless of the level of estimation risk. When $k = 4,000$, the p-values are greater than 99%. Since eliminating the exclusionary criteria does not qualitatively change the results, we continue to employ them for the remainder of our analysis.

5.3. Decomposition of the performance gains

Table 7 provides additional insights into the gains generated using the realized-volatility-based rolling estimator. The first two rows in each panel restate the results for the ex ante efficient static portfolios and for the volatility-timing strategies using the rolling estimator based on daily returns. The next four rows report the performance of the volatility-timing strategies using realized volatility as we sequentially add components of our procedure for constructing this estimator. The final two rows show the results using only the variance or covariance elements of the realized-volatility-based estimator, with the other elements replaced by those from the daily-returns-based estimator. All of the results are based on $k = 4,000$.

These results indicate that realized volatility performs reasonably well even without applying our rolling estimator framework. When we simply use the lagged realized covariance matrix as our estimator of $\boldsymbol{\Sigma}_t$, the minimum volatility strategy has a p-value of 0.57 relative to the optimal static portfolio and both performance fees are positive. For the maximum return strategy, the p-value increases to 0.99, but the fees are sensitive to the level of risk aversion. The strategies fare much worse when compared to those using the rolling estimator based on daily returns. The

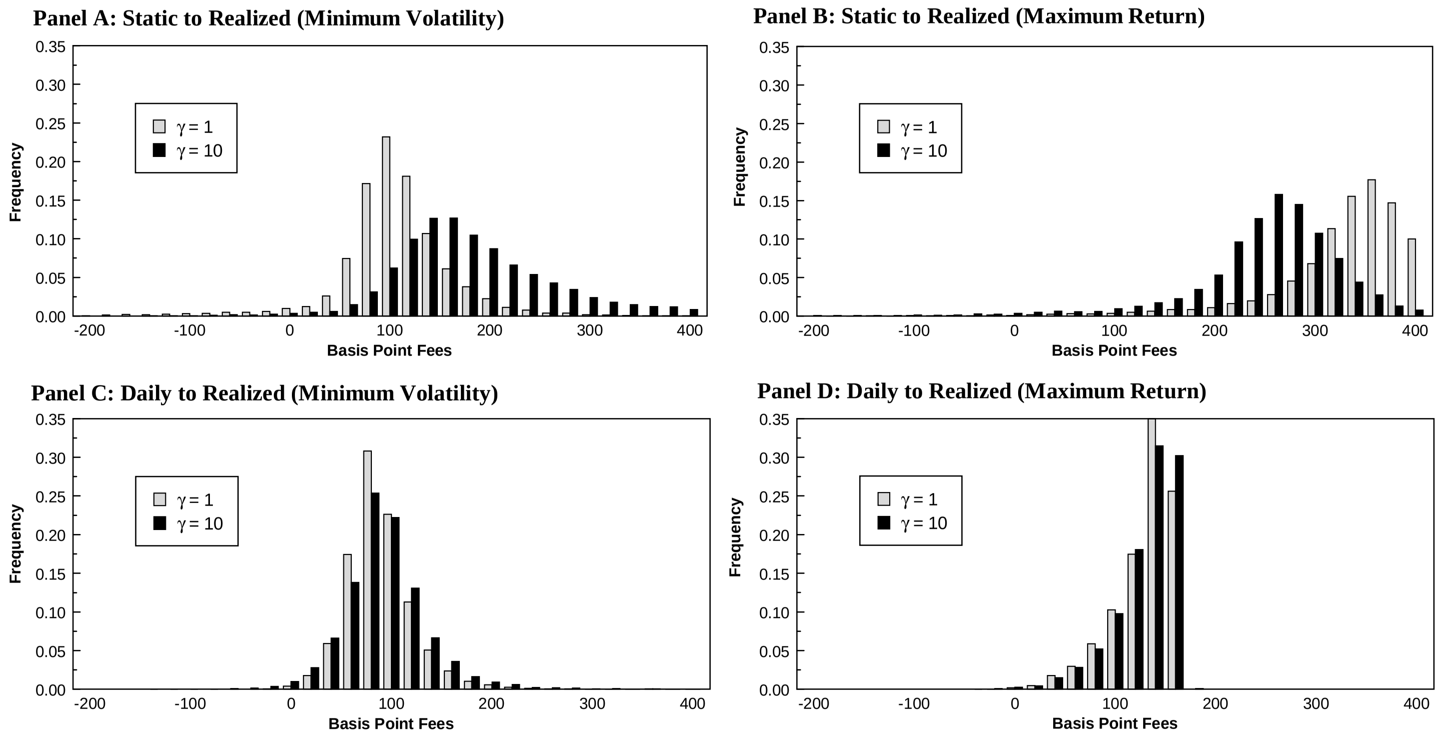

Fig. 6. Distribution of the performance fees across 10,000 bootstrap trials for the volatility-timing strategies using the realized-volatility-based estimator. In each trial, we form an artificial sample of 4,000 returns to estimate the unconditional expected returns and the unconditional covariance matrix. We then compute the next-day returns earned by the ex ante efficient static portfolios (“static”) and by the volatility-timing strategies using the rolling estimator based on either daily returns (“daily”) or realized volatility (“realized”). Finally, we estimate the fees that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch strategies. Panels A and B show the fees to switch from the static portfolios to using the realized volatility approach, and Panels C and D show the fees to switch from using daily returns to realized volatility in the rolling estimator. All of the fees are expressed as annualized basis points.

Table 6
The effect of the exclusionary criteria on the performance of the volatility-timing strategies
The table compares the performance of the volatility-timing strategies and the ex ante efficient static portfolios when the exclusionary criteria used in Table 4 are not applied. The results for each line in the table are based on 1,000 simulation trials using artificial samples of k returns to estimate the unconditional expected returns and the unconditional covariance matrix. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios ("static") and by the volatility-timing strategies using either the daily-returns-based rolling estimator ("daily") or the realized-volatility-based estimator ("realized"). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which each volatility-timing strategy earns a higher Sharpe ratio than the corresponding static portfolio, and the average annualized basis point fees (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the static portfolio to the volatility-timing strategy.

	Static			Daily						Realized					
k	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies															
1,000	6.2	11.4	0.61	6.1	10.7	0.64	0.73	3.4	119.6	6.6	10.7	0.69	0.80	50.0	169.1
2,000	8.3	11.9	0.72	8.2	11.2	0.76	0.80	2.9	92.4	8.8	11.2	0.82	0.87	67.2	157.3
3,000	8.9	12.0	0.75	8.9	11.3	0.80	0.87	9.2	90.0	9.6	11.3	0.87	0.93	83.6	165.0
4,000	9.3	12.2	0.78	9.4	11.5	0.83	0.91	13.7	97.9	10.2	11.5	0.90	0.96	96.1	181.8
5,000	9.6	12.3	0.79	9.7	11.6	0.85	0.93	14.5	96.2	10.6	11.5	0.93	0.97	102.5	185.9
10,000	10.1	12.4	0.82	10.2	11.6	0.88	0.98	17.0	98.4	11.2	11.6	0.97	0.99	117.4	203.7
Panel B: maximum return strategies															
1,000	7.7	12.5	0.61	9.0	13.1	0.69	0.83	122.4	61.2	9.9	13.1	0.76	0.85	213.9	151.6
2,000	8.9	12.4	0.72	10.6	13.0	0.81	0.91	164.1	88.1	11.7	13.1	0.89	0.94	272.3	195.5
3,000	9.3	12.3	0.75	11.2	13.0	0.86	0.95	182.8	99.5	12.3	13.0	0.95	0.96	299.0	215.0
4,000	9.5	12.2	0.78	11.5	13.0	0.88	0.97	194.5	104.6	12.8	13.0	0.98	0.98	316.1	225.6
5,000	9.7	12.2	0.79	11.8	13.0	0.90	0.99	202.0	109.9	13.0	13.0	1.00	0.99	327.7	235.2
10,000	10.0	12.1	0.82	12.2	13.0	0.94	1.00	215.3	114.3	13.6	13.0	1.04	1.00	347.7	247.0

p-values are 0.15 and 0.05 and three of the four performance fees are negative. This is not surprising, however, given the variability of the realized volatilities and correlations shown in Panel A of Figs. 2 and 3.

Rows four and five of each panel consider rolling versions of the realized-volatility-based estimator, with and without overnight returns, but before applying our bias corrections. These results indicate that using the rolling estimator reduces volatility and improves the performance of both volatility-timing strategies, especially after including overnight returns. The minimum volatility strategy now has both a higher return and lower volatility than using the rolling estimator based on daily returns; however, for the maximum return strategy, the daily-returns-based estimator is still more effective. Row six reports the results for the final bias-corrected rolling estimator based on realized volatility. Adding the bias corrections generates a large increase in the average returns for both strategies and substantially improves their overall performance. The final two rows of each panel show that the variances and covariances are of roughly equal importance in terms of their impact on economic performance.

Table 7

Decomposition of the performance gains for the realized-volatility-based estimator

The table shows the effects of the steps in our procedure for constructing the realized-volatility-based rolling estimator on volatility-timing performance. The results are based on 1,000 simulation trials using an artificial sample of 4,000 returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios ("static") and by the volatility-timing strategies using either the daily-returns-based estimator ("daily") or the realized-volatility-based estimator ("realized"). For the strategies using the realized volatility approach, we report results as we sequentially adjust our estimate of the conditional covariance matrix. First we use the lagged realized covariance matrix, next we use the unadjusted rolling estimator, then we include overnight returns in the rolling estimator, and finally we add the bias corrections. We also compute results using the "variances only" or the "covariances only" of the realized-volatility-based estimator (with the other elements based on the daily returns). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each case, the proportion of trials (p-val) in which using realized volatility earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch to using realized volatility.

				vs. static			vs. daily		
Strategy	μ	σ	SR	p-val	Δ_1	Δ_{10}	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies									
Static	9.5	12.1	0.79						
Daily	9.5	11.3	0.85	0.96	15.9	103.4			
Realized									
Lagged realized	9.5	11.8	0.81	0.57	4.3	37.8	0.15	−11.6	−65.6
Add rolling	9.4	11.3	0.83	0.96	−1.9	81.7	0.23	−17.8	−21.6
Add overnight	9.8	11.2	0.87	0.98	37.8	134.8	0.86	21.9	31.4
Add corrections	10.4	11.3	0.93	0.98	105.3	196.3	1.00	89.4	92.9
Variances only	9.9	11.2	0.89	0.98	53.0	152.0	1.00	37.2	48.7
Covariances only	10.1	11.4	0.89	0.96	67.1	143.2	0.95	51.2	39.8
Panel B: maximum return strategies									
Static	9.7	12.3	0.79						
Daily	11.8	13.0	0.91	0.99	205.0	121.2			
Realized									
Lagged realized	15.9	18.7	0.85	0.99	518.7	−388.9	0.05	313.7	−510.5
Add rolling	13.0	15.1	0.86	0.99	295.5	−53.1	0.12	90.5	−174.4
Add overnight	11.9	13.2	0.90	0.99	204.9	102.6	0.31	−0.1	−18.6
Add corrections	13.1	13.0	1.01	0.99	333.2	251.8	1.00	128.2	130.6
Variances only	12.4	13.0	0.95	0.99	256.8	180.5	1.00	51.8	59.4
Covariances only	12.6	13.1	0.96	0.98	283.9	190.8	0.98	78.9	69.6

5.4. Performance gains measured over longer horizons

Our results thus far provide strong support for the proposition that volatility timing has significant value to investors with a one-day horizon. More generally, we are interested in whether volatility timing at the daily level leads to performance gains over longer investment horizons. As discussed in Section 2, the results for a daily horizon can provide a guide to the results for longer horizons. However, the accuracy of this approach for assessing long-horizon performance depends on the time-series properties of the daily returns.

We provide direct evidence on this issue by determining the gains to daily volatility timing over horizons ranging from one week to one year. Table 8 examines the cumulative returns over nonoverlapping multi-day periods for the ex ante efficient static and volatility-timing portfolios with $k = 4,000$. The results indicate that the mean returns tend to rise and the volatilities tend to fall as the measurement horizon gets longer, leading to increases in the Sharpe ratios. However, this occurs for both the static and volatility-timing portfolios, so there is little effect on the relative performance. For each strategy, at almost every horizon, volatility timing generates a higher Sharpe ratio in over 90% of the trials and the performance fees are substantial at both levels of risk aversion. These results, therefore, confirm that our

Table 8

Performance of the volatility-timing strategies over longer horizons

The table compares the performance of the daily volatility-timing strategies and the ex ante efficient static portfolios for a range of performance measurement horizons. The results are based on 1,000 simulation trials using an artificial sample of 4,000 returns to estimate the unconditional expected returns. For each trial, we compute the realized returns over various horizons (nonoverlapping) for the ex ante efficient static portfolios ("static") and for the volatility-timing strategies using the realized-volatility-based estimator ("realized"). For each horizon, we report the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for the static and volatility-timing strategies, the proportion of trials (p-val) in which the volatility-timing strategy earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the static portfolio to the volatility-timing strategy.

		Static			Realized					
Horizon	Obs.	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies										
Daily	4,175	9.5	12.1	0.79	10.4	11.3	0.93	0.98	105.3	196.3
Weekly	834	9.5	11.6	0.82	10.4	11.2	0.94	0.97	97.2	142.2
Biweekly	417	9.5	11.4	0.84	10.4	11.0	0.96	0.97	95.2	140.7
Monthly	198	9.6	11.4	0.85	10.5	11.0	0.96	0.96	90.8	135.6
Bimonthly	98	9.7	11.3	0.87	10.5	10.5	1.02	0.98	90.7	181.9
Quarterly	65	9.8	10.9	0.92	10.6	9.8	1.10	0.98	91.8	198.0
Semiannual	32	9.8	10.6	0.95	10.7	9.3	1.17	0.97	96.7	225.0
Annual	16	9.8	10.9	0.92	10.8	10.0	1.12	0.91	114.2	213.5
Panel B: maximum return strategies										
Daily	4,175	9.7	12.3	0.79	13.1	13.0	1.01	0.99	333.2	251.8
Weekly	834	9.7	11.8	0.82	13.2	13.3	0.99	0.99	324.7	166.5
Biweekly	417	9.7	11.6	0.84	13.1	13.0	1.01	0.98	322.8	165.7
Monthly	198	9.8	11.6	0.85	13.2	13.2	1.00	0.98	319.0	143.2
Bimonthly	98	10.0	11.5	0.87	13.3	12.8	1.05	0.98	323.7	186.3
Quarterly	65	10.1	11.1	0.92	13.5	12.2	1.11	0.99	332.2	212.7
Semiannual	32	10.1	10.7	0.95	13.7	12.2	1.14	0.92	344.1	200.2
Annual	16	10.0	11.0	0.92	14.0	14.0	1.01	0.77	365.9	40.1

earlier results based on a daily horizon provide a reasonable guide to the value of volatility timing measured over longer horizons.

6. Robustness tests

In this section, we perform additional tests to assess the robustness of our results. First, we evaluate whether we should use a multivariate GARCH model instead of the daily-returns-based rolling estimator as a benchmark for measuring the incremental value of the realized volatility approach. Second, we assess the sensitivity of our results to the choice of decay rate used in implementing the rolling estimators. Third, we evaluate the impact of transaction costs on the performance of the volatility-timing strategies.

6.1. A multivariate GARCH alternative

Our results suggest that switching from daily to intradaily returns to construct the conditional covariance matrix estimates generates substantial economic value. These results, however, may overstate the value if our parsimonious daily-returns-based rolling estimator fails to adequately capture the dynamics of the conditional covariance matrix process. Using a less restrictive estimator, such as a multivariate GARCH model, may lead to different inferences regarding the incremental value of the realized volatility approach.

To investigate this issue, we fit the multivariate GARCH model in Eqs. (9) and (10) to the daily returns and then we evaluate the performance of the model relative to the daily-returns-based rolling estimator when implementing the volatility-timing strategies.[7] This is a stringent test because the fitted values of $\boldsymbol{\Sigma}_t$ from the multivariate GARCH model are based on parameter estimates obtained using the full dataset. From a statistical perspective, the multivariate GARCH model clearly fits better than the GARCH analog of the rolling estimator: the increase in the log likelihood is 105.2. The question of interest, however, is whether this statistical advantage translates into improved performance in our volatility-timing framework.

Table 9 compares the performance of the two estimators along this dimension. The *p*-values indicate the fraction of trials in which the rolling estimator generates a higher Sharpe ratio than the GARCH model. In almost every case, the *p*-values exceed 0.90. The only exceptions are for the minimum volatility strategies where the *p*-values are 0.80 and 0.85 at the two highest levels of estimation risk. Moreover, all of the performance fees for switching from the GARCH model to the rolling estimator are positive. For the minimum volatility strategies they range from 12 to 38 basis points and for the maximum return strategies they range from 25 to 147 basis points.

While it may seem surprising that the more restrictive rolling estimator delivers the better performance, this is consistent with evidence reported in FKO (2001). They

[7] We restrict the $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices to be diagonal to avoid overparameterizing the model.

Table 9
Performance of the daily-returns-based estimator vs. GARCH
The table compares the performance of the volatility-timing strategies using the daily-returns-based rolling estimator of conditional volatility versus using the in-sample fitted values from a multivariate GARCH model. The results for each line in the table are based on 1,000 simulation trials using an artificial sample of k returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the volatility-timing strategies using either the GARCH model ("GARCH") or the daily-returns-based estimator ("daily"). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which the daily-returns-based estimator earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from using GARCH to using the rolling estimator.

	GARCH			Daily					
k	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies									
1,000	7.0	9.1	0.77	7.2	9.1	0.79	0.80	19.5	11.9
2,000	8.1	10.3	0.80	8.4	10.4	0.82	0.85	24.2	13.5
3,000	8.8	10.9	0.81	9.1	11.0	0.83	0.90	28.0	16.7
4,000	9.2	11.2	0.83	9.5	11.3	0.85	0.91	30.8	18.1
5,000	9.2	11.1	0.84	9.6	11.2	0.86	0.94	32.0	20.2
10,000	9.8	11.4	0.86	10.2	11.5	0.89	0.99	37.8	24.3
Panel B: maximum return strategies									
1,000	9.6	12.1	0.79	11.0	13.1	0.85	0.99	131.0	24.8
2,000	9.9	12.1	0.82	11.4	13.1	0.87	1.00	135.9	30.3
3,000	10.2	12.1	0.84	11.7	13.1	0.89	1.00	138.7	33.6
4,000	10.3	12.1	0.85	11.8	13.0	0.91	1.00	140.8	36.1
5,000	10.4	12.1	0.86	12.0	13.0	0.92	1.00	142.0	37.7
10,000	10.8	12.1	0.89	12.3	13.0	0.95	1.00	146.6	43.1

find that volatility timing adds the most value using estimates of $\boldsymbol{\Sigma}_t$ that are much smoother than is optimal under a statistical goodness-of-fit criterion. Apparently the fitted values from the multivariate GARCH model exhibit too much variability when viewed through the lens of the volatility-timing strategies.

6.2. The effect of the decay rate

The next robustness issue we consider is the choice of decay rate used to implement the two rolling estimators. Our previous results are based on the decay rates that provide the best statistical fit for the GARCH analog of the rolling estimators. However, as we found with the multivariate GARCH model, using a statistical criteria may be suboptimal in terms of implementing the volatility-timing strategies. This has two potential implications for our results. First, if the decay rate for the daily-returns-based estimator is suboptimal, our results may overstate the incremental value of using the realized volatility approach. Second, if the decay rate for the realized-volatility-based estimator is suboptimal, our results may understate the overall value of volatility-timing using realized volatility.

Table 10

The effect of the decay rate on the performance of the daily-returns-based estimator

The table shows the effect of the decay rate (α) used in the daily-returns-based rolling estimator on the performance of the volatility-timing strategies. The results are based on 1,000 simulation trials using an artificial sample of 4,000 returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios ("static") and for the volatility-timing strategies using either the daily-returns-based estimator ("daily") or the realized-volatility-based estimator ("realized"). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which using daily returns earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch to using the daily-returns-based estimator.

					vs. static			vs. realized		
Strategy	α	μ	σ	SR	p-val	Δ_1	Δ_{10}	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies										
Static	—	9.5	12.1	0.79						
Realized	0.064	10.4	11.3	0.93	0.98	105.3	196.3			
Daily	0.001	9.8	12.0	0.82	0.92	28.5	36.7	0.02	−76.8	−159.6
	0.005	9.9	11.7	0.85	0.96	47.9	95.2	0.02	−57.4	−101.1
	0.010	9.7	11.5	0.85	0.97	29.4	95.5	0.02	−75.9	−100.8
	0.020	9.6	11.4	0.85	0.97	18.4	100.3	0.01	−86.9	−96.0
	0.040	9.5	11.3	0.85	0.95	12.6	99.9	0.00	−92.7	−96.4
	0.060	9.4	11.4	0.83	0.90	−1.0	78.1	0.00	−106.3	−118.2
	0.080	9.2	11.5	0.81	0.72	−17.8	48.0	0.00	−123.1	−148.2
	0.100	9.1	11.6	0.79	0.51	−34.3	16.1	0.00	−139.6	−180.1
Panel B: maximum return strategies										
Static	—	9.7	12.3	0.79						
Realized	0.064	13.1	13.0	1.01	0.99	333.2	251.8			
Daily	0.001	9.9	11.9	0.84	0.98	29.3	77.2	0.01	−303.9	−174.5
	0.005	11.4	12.3	0.92	0.99	166.7	168.9	0.01	−166.5	−82.8
	0.010	11.5	12.5	0.92	0.99	178.0	160.9	0.01	−155.2	−90.8
	0.020	11.7	12.7	0.92	0.99	195.2	146.5	0.01	−138.0	−105.3
	0.040	11.9	13.3	0.89	0.98	208.1	92.0	0.00	−125.1	−159.8
	0.060	11.9	13.9	0.86	0.96	203.3	12.6	0.00	−129.9	−239.2
	0.080	11.9	14.5	0.82	0.90	191.3	−79.7	0.00	−141.9	−331.4
	0.100	11.9	15.2	0.78	0.39	177.7	−178.7	0.00	−155.5	−430.3

6.2.1. Sensitivity analysis for the daily-returns-based estimator

Table 10 shows how our results vary with the decay rate used to implement the daily-returns-based estimator. The first two rows in each panel restate the results for the ex ante efficient static portfolios and for the realized-volatility-based estimator with $\alpha = 0.064$. The results in the remaining rows show that using a decay rate of about 0.005 for the daily-returns-based estimator maximizes the overall performance of the volatility-timing strategies. With this choice, the p-values relative to the static portfolios are 0.96 and 0.99 and the performance fees are 48 and 95 basis points for the minimum volatility strategy and 167 and 169 basis points for the maximum return strategy. However, using the realized-volatility-based estimator still leads to better performance. The *p*-values indicate that switching from realized volatility to the daily-returns-based estimator rarely yields a higher Sharpe ratio and all of the performance fees are negative, independent of the choice of decay rate.

Fig. 7 shows the distribution of the performance fees across the bootstrap trials using a decay rate of 0.005 for the daily-returns-based estimator. Panels A and B show the fees for switching from the ex ante efficient static portfolios to the volatility-timing strategies using the daily-returns-based estimator. Panels C and D show the fees for switching from the daily-returns-based estimator to the realized-volatility-based estimator. In every case the majority of the distribution lies above zero. Moreover, the distribution in Panel A is roughly comparable to that in Panel C, and the same is true for the distributions in Panels B and D. From this we conclude that, even if we implement the daily-returns-based estimator using the most effective decay rate, switching to the realized-volatility-based estimator approximately doubles the economic value of volatility timing.

6.2.2. *Sensitivity analysis for the realized-volatility-based estimator*

In Table 11, we show how our results vary with the decay rate used to implement the realized-volatility-based estimator. The first two rows in each panel restate the results for the ex ante efficient static portfolios and for the daily-returns-based estimator with $\alpha = 0.005$. The remaining rows show that as we vary the decay rate used for the realized-volatility-based estimator from 0.020 to 0.175, the performance of the volatility-timing strategies is not greatly affected. In almost every case, the *p*-values with respect to both the ex ante efficient static portfolios and the daily-returns-based estimator are greater than 0.90 and switching to the realized-volatility-based estimator generates large performance fees.

Based on these results, we can provide some conservative estimates of the economic value of volatility timing using realized volatility. The overall value of volatility-timing using realized volatility is approximately 100 to 200 basis points for the minimum volatility strategy and 250 to 350 basis points for the maximum return strategy. The incremental value of using realized volatility, compared to the daily-returns-based estimator, is approximately 50 to 100 basis points for the minimum volatility strategy and 80 to 200 basis points for the maximum return strategy. Thus, it seems clear that the realized volatility approach has considerable economic value.

6.3. *Transaction costs*

Evaluating the effect of transaction costs provides another perspective on the economic significance of our results. To get a sense of a reasonable level of transaction costs, consider the S&P 500 futures contract. Fleming et al. (1996) estimate that the roundtrip commissions and fees for large institutions are about \$6.00 per contract. If we assume, conservatively, that the bid-ask spread averages 0.10 (\$50 per contract) and use the average futures price during our sample (551.39), we obtain an average one-way transaction cost of \$28.00 on a contract size

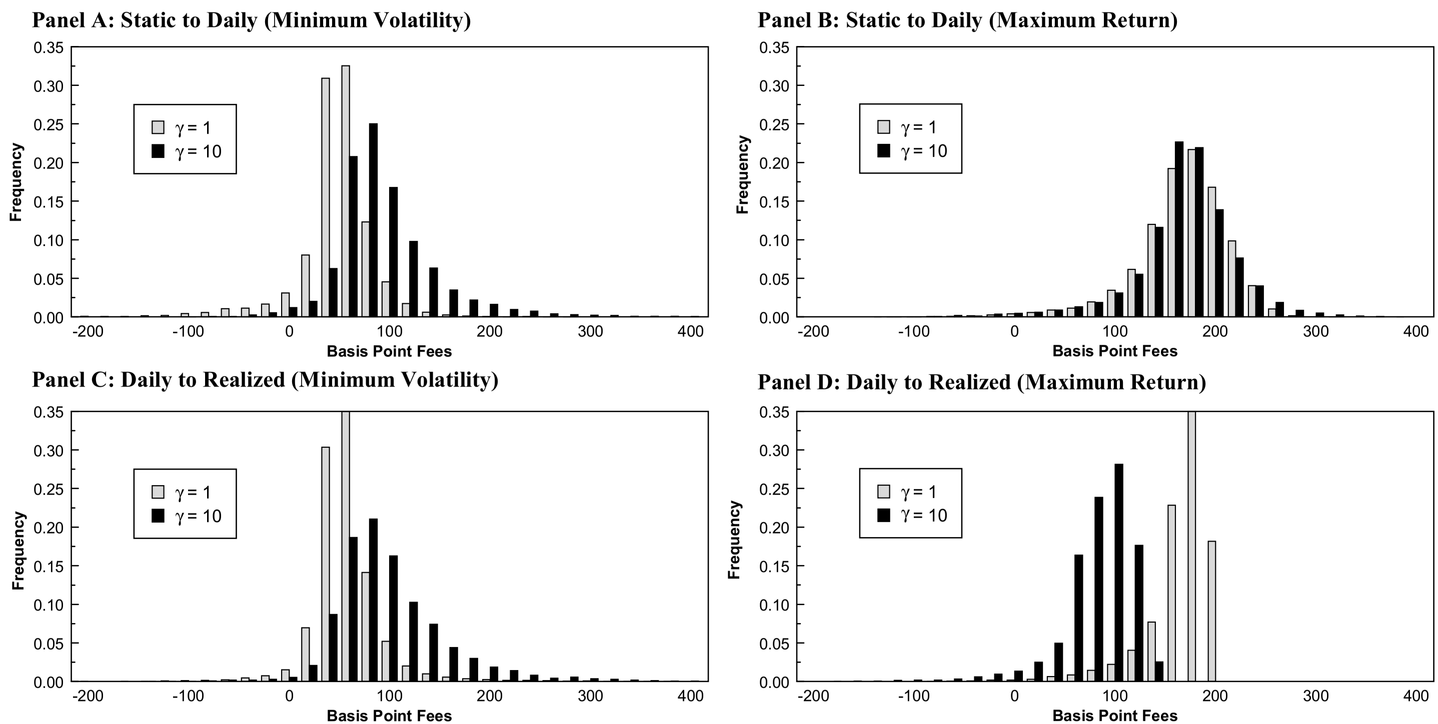

Fig. 7. Distribution of the performance fees across 10,000 bootstrap trials using $\alpha = 0.005$ for the daily-returns-based estimator. In each trial, we form an artificial sample of 4,000 returns to estimate the unconditional expected returns and the unconditional covariance matrix. We then compute the next-day returns earned by the ex ante efficient static portfolios ("static") and by the volatility-timing strategies using the rolling estimator based on either daily returns ("daily") or realized volatility ("realized"). Finally, we estimate the fees that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch strategies. Panels A and B show the fees to switch from the static portfolios to using the daily-returns-based estimator, and Panels C and D show the fees to switch from using daily returns to realized volatility in the rolling estimator. All of the fees are expressed as annualized basis points.

Table 11
The effect of the decay rate on the performance of the realized-volatility-based estimator
The table shows the effect of the decay rate (α) used in the realized-volatility-based rolling estimator on the performance of the volatility-timing strategies. The results are based on 1,000 simulation trials using an artificial sample of 4,000 returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios (“static”) and for the volatility-timing strategies using either the daily-returns-based estimator (“daily”) or the realized-volatility-based estimator (“realized”). The table reports the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for each strategy, the proportion of trials (p-val) in which the realized-volatility-based estimator earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch to using realized volatility.

					vs. static			vs. daily		
Strategy	α	μ	σ	SR	p-val	Δ_1	Δ_{10}	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies										
Static	—	9.5	12.1	0.79						
Daily	0.005	9.9	11.7	0.85	0.96	47.9	95.2			
Realized	0.020	9.9	11.4	0.88	0.99	52.6	134.6	0.94	4.8	39.4
	0.040	10.2	11.3	0.91	0.98	84.5	176.8	0.99	36.6	81.7
	0.060	10.4	11.3	0.93	0.98	102.4	194.2	0.99	54.5	99.0
	0.080	10.5	11.3	0.94	0.98	114.1	200.3	0.98	66.2	105.1
	0.100	10.6	11.4	0.94	0.97	124.5	200.3	0.96	76.6	105.1
	0.125	10.7	11.6	0.92	0.94	122.5	175.1	0.84	74.7	79.9
	0.150	10.7	12.2	0.89	0.76	123.0	106.9	0.69	75.1	11.7
	0.175	10.7	12.7	0.86	0.68	117.4	37.9	0.63	69.5	−57.4
Panel B: maximum return strategies										
Static	—	9.7	12.3	0.79						
Daily	0.005	11.4	12.3	0.92	0.99	166.7	168.9			
Realized	0.020	12.4	13.1	0.94	0.99	255.8	162.0	0.90	89.1	−7.0
	0.040	12.8	13.0	0.98	0.99	296.7	217.0	0.99	130.0	48.0
	0.060	13.1	13.0	1.00	0.99	327.6	247.4	0.99	160.9	78.4
	0.080	13.3	13.1	1.02	0.99	349.7	261.6	0.99	183.0	92.6
	0.100	13.6	13.2	1.03	0.99	374.8	273.8	0.97	208.0	104.8
	0.125	13.6	13.4	1.02	0.98	379.8	253.8	0.95	213.1	84.8
	0.150	13.9	14.1	0.98	0.93	391.7	173.1	0.72	225.0	4.0
	0.175	14.0	14.9	0.94	0.79	391.0	66.8	0.65	224.3	−102.4

\$551.39 $\times$ 500, or 2.56% annualized.[8] This means that if, on average, we trade one contract every day for a year, transaction costs would reduce our returns by 256 basis points.

Table 12 illustrates the effect of imposing different levels of proportional transaction costs on the performance of the ex ante efficient static portfolios and

[8] This assumes a contract size of 500 times the index level. The contract size for S&P 500 futures changed from 500 to 250 times the index level in November 1997.

Table 12
The effect of transaction costs on the performance of the realized-volatility-based estimator
The table compares the performance of the volatility-timing strategies using the realized-volatility-based estimator to that of the ex ante efficient static portfolios under various levels of transaction costs. The results are based on 1,000 simulation trials using an artificial sample of 4,000 returns to estimate the unconditional expected returns. For each trial, we compute the next-day returns earned by the ex ante efficient static portfolios ("static") and for the volatility-timing strategies using the realized-volatility-based estimator ("realized"). We impose proportional transaction costs on all trades, either rebalancing or rolling into the next contract month. These "Pct. Costs" are expressed in the table as the percentage one-way cost of trading one futures contract every day for a year. For each level of transaction costs, we report the average annualized mean return (μ), annualized volatility (σ), and Sharpe ratio (SR) for the static and volatility-timing strategies, the proportion of trials (p-val) in which the volatility-timing strategy earns a higher Sharpe ratio, and the average annualized basis point fees (Δ_γ) that an investor with quadratic utility and constant relative risk aversion of γ would be willing to pay to switch from the static portfolio to the volatility-timing strategy.

	Static			Realized					
Pct. Costs	μ	σ	SR	μ	σ	SR	p-val	Δ_1	Δ_{10}
Panel A: minimum volatility strategies									
0.0%	9.5	12.1	0.79	10.4	11.3	0.93	0.98	105.3	196.3
2.5%	9.4	12.1	0.78	10.2	11.3	0.91	0.98	92.9	184.0
5.0%	9.2	12.1	0.77	9.9	11.3	0.89	0.97	80.5	171.6
7.5%	9.1	12.1	0.76	9.7	11.3	0.87	0.97	68.1	159.3
10.0%	9.0	12.1	0.75	9.5	11.3	0.84	0.96	55.7	147.0
12.5%	8.9	12.1	0.74	9.2	11.3	0.82	0.96	43.3	134.7
15.0%	8.7	12.1	0.73	9.0	11.3	0.80	0.94	30.9	122.4
17.5%	8.6	12.1	0.72	8.7	11.3	0.78	0.92	18.6	110.1
20.0%	8.5	12.1	0.71	8.5	11.3	0.76	0.90	6.2	97.7
Panel B: maximum return strategies									
0.0%	9.7	12.3	0.79	13.1	13.0	1.01	0.99	333.2	251.8
2.5%	9.6	12.3	0.78	12.8	13.0	0.98	0.99	312.4	231.0
5.0%	9.5	12.3	0.77	12.5	13.0	0.96	0.99	291.6	210.3
7.5%	9.3	12.3	0.76	12.1	13.0	0.93	0.98	270.8	189.6
10.0%	9.2	12.3	0.75	11.8	13.0	0.91	0.98	250.0	168.8
12.5%	9.1	12.3	0.74	11.5	13.0	0.88	0.98	229.2	148.0
15.0%	9.0	12.3	0.73	11.1	13.0	0.85	0.97	208.4	127.3
17.5%	8.8	12.3	0.72	10.8	13.0	0.83	0.96	187.6	106.5
20.0%	8.7	12.3	0.71	10.5	13.0	0.80	0.95	166.8	85.7

the volatility-timing strategies. We impose the transaction costs on every trade, including those required to establish the initial position, implement the daily rebalancing, roll into each subsequent contract month, and liquidate the position at the end of the sample. For the volatility-timing strategies, daily rebalancing is needed to track the time variation in the portfolio weights and, for the static portfolios, it is needed to maintain a constant weight in each asset. We impose the transaction costs by subtracting the appropriate percentage each day from the daily portfolio returns.

As expected, transaction costs have a larger impact on the volatility-timing strategies than on the ex ante efficient static portfolios. Transaction costs of 10%, for example, would reduce the average return on the volatility-timing strategies by 90 and 130 basis points compared to only 50 basis points for the static portfolios. Nonetheless, even with this level of transaction costs, both of the volatility-timing strategies outperform the static portfolios by a substantial margin. The Sharpe ratios are higher in over 95% of the trials, and the performance fees are 56 and 147 for the minimum volatility strategy and 250 and 169 for the maximum return strategy. Thus, transaction costs would need to be much larger than our estimate of 2.56% in order to offset the gains to volatility timing using realized volatility.

A separate question is whether the gains may be offset by market-impact costs. To examine this, suppose we implement the maximum return strategy using a constant $100 million value for the portfolio. In our bootstrap experiment, managing such a position requires, on average, daily trade sizes of 16, 40, and 93 contracts for stock, bond, and gold futures, respectively. For comparison, the average daily trading volumes in these markets during our sample period are 69,000 for stocks, 290,000 for bonds, and 31,000 for gold. Thus, it seems unlikely that market-impact costs would be large enough to offset the observed performance differential.

7. Conclusions

Prior research has documented many statistical benefits of the realized volatility approach. We find that, with regard to investment decisions, the approach yields substantial economic benefits as well. In particular, an investor implementing a volatility-timing strategy would be willing to pay on the order of 50 to 200 basis points per year to switch from a daily-returns-based estimator of the conditional covariance matrix to an estimator based on realized volatility. Relative to an ex ante efficient static portfolio, the overall value of volatility timing using realized volatility is approximately double this magnitude. Moreover, we find that these benefits are not restricted to short-horizon investors. The cumulative gains to volatility timing at the daily level are substantial for performance measurement horizons as long as a year.

The results of our analysis point to some interesting directions for future research. One possibility is to relax the assumption that our investor treats expected returns as constant. We could then attempt to relate the predictable variation in monthly returns to persistence in daily expected returns which in turn is related to persistence in conditional covariances. Establishing a clear linkage between conditional expected returns and covariance-based measures of systematic risk has proven difficult in the past. Perhaps using the realized volatility approach will be more successful.

Another potential avenue is an in-depth analysis of the statistical performance of realized-volatility-based estimators of the conditional covariance matrix. ABDL (2003) show that modeling the time series of realized volatilities as a simple autogressive process produces volatility forecasts that outperform those obtained by fitting standard GARCH models. Given our results, it would be interesting to compare the performance of the ABDL (2003) specification to that of GARCH models formulated directly in terms of the lagged realized volatilities.

References

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica, forthcoming.

Andreou, E., Ghysels, E., 2002. Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. Journal of Business and Economic Statistics 20, 363–376.

Areal, N., Taylor, S.J., 2002. The realized volatility of FTSE-100 futures prices. Journal of Futures Markets 22, 627–648.

Bai, X., Russell, J.R., Tiao, G.C., 2001. Beyond Merton's utopia (I): effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. Unpublished working paper, University of Chicago.

Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian OU based models and some of their uses in financial economics. Journal of the Royal Statistical Society, Series B 63, 167–241.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. Econometrica 41, 135–155.

Ebens, H., 1999. Realized stock index volatility. Unpublished working paper #420. Department of Economics, Johns Hopkins University.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1–26.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Fleming, J., Ostdiek, B., Whaley, R.E., 1996. Trading costs and the relative rates of price discovery in the stock, futures, and option markets. Journal of Futures Markets 16, 353–387.

Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

Hall, P., Horowitz, J., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891–916.

Hsieh, D.A., 1991. Chaos and nonlinear dynamics: application to financial markets. Journal of Finance 46, 1839–1877.

Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. Journal of Financial Economics 8, 323–361.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Schwert, G.W., 1989. Does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester, UK.