Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Empirical Finance 10 (2003) 533–558
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

Measuring and modeling systematic risk in factor pricing models using high-frequency data

Tim Bollerslev[a,b,*], Benjamin Y.B. Zhang[a,1]

[a] *Department of Economics, Duke University, Durham, NC 27708, USA*
[b] *NBER, USA*

Received 28 June 2001; accepted 7 November 2002

Abstract

This paper demonstrates how high-frequency data may be used in more effectively measuring and modeling the systematic risk(s) in factor pricing models. Based on a 7-year sample of continuously recorded US equity transactions, we find that simple and easy-to-implement time series forecast for the high-frequency-based factor loadings in the three-factor Fama–French model gives rise to more accurate factor representations and improved asset pricing predictions when compared to the conventional monthly rolling regression-based estimates traditionally employed in the literature, in turn resulting in more efficient ex post mean-variance portfolios. As such, the methodology proposed in the paper holds the promise for important new insights concerning actual real-world investment decisions and practical situations involving risk management.

JEL classification: C14; C33; G12
Keywords: High-frequency data; Factor pricing models; Realized loadings; Fama–French factors

1. Introduction

The advent of high-frequency transactions, or tick-by-tick, prices for a host of different financial instruments have spurred a renewed interest into improved measure-

* Corresponding author. Department of Economics, Duke University, Durham, NC 27708-0097, USA. Tel.: +1-919-660-1846; fax: +1-919-684-6974.
E-mail addresses: boller@econ.duke.edu (T. Bollerslev), yz6@acpub.duke.edu (B.Y.B. Zhang).
[1] Tel.: +1-919-660-1878.

0927-5398/03/$ - see front matter
doi:10.1016/S0927-5398(03)00004-5

ments of financial market volatility.[2] The basic intuition is straightforward. By summing high-frequency squared returns, it is possible to more accurately measure the ex post *realized* volatility over a fixed time interval.[3] Doing so, a number of new and interesting empirical regularities have emerged. In particular, consistent with earlier results related to the foreign exchange market in Andersen et al. (2001a), the study by Andersen et al. (2001b) finds that the realized daily volatilities for the 30 stocks in the Dow Jones Industrial Average are approximately log-normally distributed with apparent long-memory dependencies, while the returns standardized by their realized volatilities are close to Gaussian.

Meanwhile, from a pricing perspective only the systematic, or the nondiversifiable, portion of the volatility should be priced. Asset pricing models allowing for time-varying CAPM betas or multi-factor loadings based on ARCH type formulations that explicitly incorporate conditional information variables have previously been estimated by Bollerslev et al. (1988), Ball and Kothari (1989), and Braun et al. (1995) among others. These studies generally find that explicitly allowing for temporal variation in the factor loading results in statistically significant and economically important improvements in the performance of the models.

Building on these ideas, the present paper demonstrates how the systematic risk in standard multi-factor asset pricing models may be assessed more accurately through the use of high-frequency data and realized volatility measures.[4] For concreteness, we focus on the Fama and French (1993) three-factor model, but the methodology is general and could be applied in the measurement of standard CAPM betas, or factor loadings in other ICAPM or APT type models based on factor replicating portfolios.

The three-factor Fama–French model was originally motivated by the empirical findings in Fama and French (1992, 1993) of anomalous high historical returns for small firms with high book-to-market values. In addition to the excess return on the market portfolio, MKT, the systematic risks in the model are captured by the two additional SML and HML zero-cost factor mimicking portfolio returns. As such, the model represents a generalization of the standard CAPM model. The SMB portfolio returns are defined by a zero net investment long in small and short in big capitalization stocks, while the HML factor is given by a portfolio long in high book-to-market, or value stocks, and short in low book-to-market, or growth stocks. The empirical validity of the Fama–French model, and the economic justification behind the three factors, have been

[2] An incomplete list of these studies include Andersen et al. (2001a, 2003b), Andreou and Ghysels (2002), Bai et al. (2000), Barndorff-Nielson and Shephard (2001, 2002a,b), Dacorogna et al. (2001), Meddahi (2002), and Taylor and Xu (1997); see also Andersen et al. (2003a) for a recent survey of the literature.

[3] This is distinctly different from the earlier contributions by Merton (1980) and Nelson (1990), establishing that the *point-in-time*, or spot, volatility in continuous time diffusions may be estimated arbitrarily well by appropriately filtering sufficiently high-frequency returns. The practical implementation of such filters is severely hampered by market microstructure frictions.

[4] The idea of utilizing high-frequency based *realized* volatility measures in calculating realized CAPM betas and loadings in multi-factor pricing models was initially suggested in the concluding remarks by Andersen et al. (2001b). However, the actual empirical analysis in that paper does not explore such measures, let alone any pricing implications.

the subject of an intensive debate in the asset pricing literature over the past decade.[5] However, due to the limited 7-year time-span of high-frequency data available for our empirical analysis, the present paper should not be viewed as a direct contribution to that literature per se. Rather, the paper simply illustrates the feasibility of using the high-frequency data in more effectively measuring and modeling the temporal variation in the factor loadings in a realistic empirical setting.

In this regard, the recent empirical evidence in Ferson and Harvey (1999) based on various economy-wide variables intended to capture changes in the investment opportunity set, strongly suggest that the factor loadings in the Fama–French model are indeed time-varying. Similarly, Lettau and Ludvigson (2001) report strong evidence in favor of time-varying factor loadings when conditioning on the aggregate consumption–wealth ratio. However, exactly which conditioning variables to include when implementing a particular model are not clear.[6] Meanwhile, as long as the factors fully characterize the investment opportunity set, the use of higher frequency data and corresponding realized factor loadings provide a simple way to more effectively deal with this conundrum.

To facilitate comparisons with the extant empirical literature, we focus on a 1-month return horizon and the pricing of the 25 size and book-to-market sorted portfolios first analyzed by Fama and French (1993). The pricing of these same 25 portfolios has been the subject of numerous subsequent empirical studies (e.g., Daniel and Titman, 1997; Fama and French, 1996; Ferson and Harvey, 1999; Harvey and Siddique, 2000; Hodrick and Zhang, 2001; Lettau and Ludvigson, 2001, among many others), and as such they form a natural test case for assessing the usefulness of the high-frequency data within a common framework. Although direct comparisons with other existing methods is limited by our relatively short 7-year time span of high-frequency data, we find that the return predictions utilizing the high-frequency realized loadings systematically outperform the conventional monthly rolling regression estimates traditionally employed in the literature in terms of a series of ex post statistical tests and implied mean-variance portfolio decisions.

The plan for the rest of the paper is as follows. Section 2 starts with a brief discussion of a general factor pricing model and the theory used in formally justifying our notion of realized factor loadings. Section 3 contains a discussion of the high-frequency data and procedures employed in the construction of the portfolio returns and realized factor loadings, including a simple adjustment procedure to account for

[5] For instance, in a series of subsequent papers Fama and French (1995, 1996, 1998) and Davis et al. (2000) have argued that by including the additional two risk factors the size and value premium largely disappears. Direct criticism of these empirical findings centering on survivorship and data-snooping biases include Berk (2000), Daniel and Titman (1997), Kothari et al. (1995), and Wang (2000). Possible theoretical linkages between the Fama–French factors and the underlying economic state variables have been pursued by Brennan et al. (2002), Gomes et al. (in press), and Liew and Vassalou (2000); see also Cochrane (2001) for a recent textbook discussion.

[6] As discussed by Hansen and Richard (1987) in the context of a general dynamic asset pricing model, the use of too coarse an information set may formally invalidate the model. The results in Ghysels (1998) also call into question the usefulness of poorly estimated time-varying factor loadings when empirically implementing factor pricing models.

nonsynchronous trading effects. This section also presents standard summary statistic for the monthly realized portfolio returns and factor loadings for the 25 test portfolios and the three-factor Fama–French model over the 7-year (1993–1999) sample period. Section 4 details the accuracy of several different procedures for forecasting the monthly factor loadings, followed by a comparison of the resulting asset prices and mean-variance efficient portfolios. Section 5 concludes with some suggestions for future research.

2. Theory

The true exposure of a particular financial asset to a specific source of systematic risk is generally latent. However, under suitable assumptions about the underlying return generating process, the corresponding factor loading(s) may in theory be estimated arbitrarily well through the use of sufficiently finely sampled high-frequency data. This section briefly formalizes this idea and also discusses the setup employed in the actual empirical implementation.

2.1. Factor pricing models and realized loadings

Let $r_{t+1} \equiv p(t+1) - p(t)$ denote an $N \times 1$ vector of discretely sampled continuously compounded one-period returns from time t to $t+1$. Following common practice in the empirical asset pricing literature, assume that the underlying discrete-time return generating process takes the form of a k-factor model,

$$r_{t+1} = \mu_{t+1|t} + \beta_{t+1|t}{}' f_{t+1} + \mathbf{\Omega}_{t+1|t}^{1/2} \varepsilon_{t+1}, \qquad t = 0, 1, \ldots, T-1, \tag{1}$$

where the $k \times 1$ vector of excess factor returns or systematic risks, $f_{t+1} \equiv v(t+1) - v(t)$, and the $N \times 1$ vector of idiosyncratic risks, ε_{t+1}, are conditionally uncorrelated, and otherwise satisfy $E_t(\varepsilon_{t+1}) = 0$, and $E_t(\varepsilon_{t+1}\varepsilon_{t+1}{}') = I$. Note that the $N \times N$ conditional covariance matrix, $\mathbf{\Omega}_{t+1|t}$, is allowed to depend nontrivially on the time t information set. The factor loadings are then formally defined by,

$$\beta_{t+1|t} \equiv \mathrm{Cov}_t(f_{t+1}, f_{t+1})^{-1} \times \mathrm{Cov}_t(f_{t+1}, r_{t+1}), \qquad t = 0, 1, \ldots, T-1. \tag{2}$$

Also, the absence of arbitrage implies that (see, e.g., Cochrane, 2001; Hodrick and Zhang, 2001),

$$E_t(r_{t+1}) = \boldsymbol{\iota} \times r_{ft} + \beta_{t+1|t}{}' \Lambda_{t+1|t}, \qquad t = 0, 1, \ldots, T-1, \tag{3}$$

where $\Lambda_{t+1|t} \equiv E_t(f_{t+1})$, $\boldsymbol{\iota}$ denotes an $N \times 1$ vector of ones, and r_{ft} refers to the one-period risk-free rate.

Now suppose that finer sampled h-period returns, $r_{t+ih,h} \equiv p(t+ih) - p(t+(i-1)h)$, $i=1, 2, \ldots, 1/h$, are available, where for notational simplicity we assume that $1/h$ is an integer. Moreover, assume that the intraperiod returns adhere to the same factor structure as in Eq. (1) with constant intraperiod drift, factor loadings, and covariance matrix. That is,

$$r_{t+ih,h} = \mu_{t+1|t} \times h + \beta_{t+1|t}{}' f_{t+ih,h} + \mathbf{\Omega}_{t+1|t}^{1/2} \varepsilon_{t+ih,h}, \qquad t = 0, 1, \ldots, T-1,$$
$$i = 1, 2, \ldots, 1/h, \tag{4}$$

where $f_{t+ih,h} \equiv v(t+ih) - v(t+(i-1)h)$, $\varepsilon_{t+1} \equiv \varepsilon_{t+h,h} + \varepsilon_{t+2h,h} \ldots + \varepsilon_{t+1,h}$, and $E_{t+(i-1)h}(\varepsilon_{t+ih,h}) = 0$.[7] The temporally aggregated one-period returns defined by Eq. (4), $r_{t+h,h} + r_{t+2h,h} \ldots r_{t+1,h}$, is obviously identically equal to r_{t+1} defined by Eq. (1). As previously noted, the most popular discrete-time asset pricing models employed in the literature may be represented in this form. In particular, the standard CAPM model obtains for $k=1$ and f_{t+1} equal to the instantaneous return on the market portfolio, while the Fama–French model underlying the empirical analysis below corresponds to $k=3$ with the addition of the returns on the size and book-to-market factor representing portfolios.

The conditional betas, or factor loadings, in Eqs. (1) and (4), are, of course, not directly observable. However, the elements in $\beta_{t+1|t}$ may in theory be estimated arbitrarily well through the use of increasingly finer sampled high-frequency returns. Specifically, let

$$[v, v]_{t+1} = \sum_{i=1,\ldots,1/h} f_{t+ih,h} f_{t+ih,h}{}', \qquad t = 0, 1, \ldots, T-1, \tag{5}$$

and

$$[v, p]_{t+1} = \sum_{i=1,\ldots,1/h} f_{t+ih,h} r_{t+ih,h}{}', \qquad t = 0, 1, \ldots, T-1. \tag{6}$$

The OLS estimate obtain from the multiple regression of $r_{t+ih,h}$ on $f_{t+ih,h}$,

$$\hat{\beta}_{t+1} = ([v, v]_{t+1})^{-1} [v, p]_{t+1}, \qquad t = 0, 1, \ldots, T-1, \tag{7}$$

then converges in probability to the $k \times N$ $\beta_{t+1|t}$ matrix of conditional factor loadings for $h \to 0$ under standard assumptions. Note, the jth column in $\hat{\beta}_{t+1}$ is simply equal to the coefficients in the standard univariate regression of the intraperiod returns for the jth asset on the corresponding intraperiod returns for the k factor representing portfolios. Following the recent literature on quadratic variation and related volatility measures (e.g., Andersen

[7] The general continuous time analogue of the factor representation in Eq. (1) would have μ_t, β_t and Ω_t continuously varying within the period. The assumption of constant within period factor loadings directly parallels the (implicit) assumption underlying the monthly rolling regression-based estimates for beta traditionally employed in the empirical asset pricing literature (e.g., Fama and MacBeth, 1973).

et al., 2002a, 2001a, in press; Barndorff-Nielson and Shephard, 2002a,b; Comte and Renault, 1998; Meddahi, 2002), we will refer to these one-period estimates constructed from the high-frequency intraperiod returns as *realized* factor loadings.[8]

Although the regression estimates in Eq. (7) converges to the true (latent) factor loadings for infinitesimally small h, data limitations invariably put an upper bound on the highest empirically feasible sampling frequency. The next section discusses these issues as they relate to the actual stock return data employed in the empirical analysis.

3. Data and measurements

We begin this section by a brief discussion of the data sources and pertinent market microstructure issues involved in the construction of the high-frequency-based factor loadings. We then turn to a discussion of the resulting monthly realized factor loadings and monthly returns for the three-factor Fama–French model and the 25 select size and book-to-market sorted portfolios.

3.1. High-frequency data and nonsynchronous trading

The data sources include the Trade And Quote (TAQ), Center for Research in Security Prices (CRSP), Compustat, and CRSP/Compustat merged datasets. All of the databases are maintained by the Wharton Research and Data Services (WRDS). The TAQ database is issued by the New York Stock Exchange, and contains intraday trades and quotes for all securities listed on the NYSE, AMEX, and NASDAQ stock markets. We rely on the consolidated trade database in extracting individual stock prices. The availability of the TAQ database limits our empirical analysis to the January 1993 through December 1999 time period.

The consistency results in the previous section dictates that the returns be measures over infinitesimally small time periods. Of course, in actuality financial prices are only available at irregularly spaced discrete time intervals. Moreover, there is a clear tendency for the observed prices to cluster at discrete price points (see, e.g., Bollerslev and Melvin, 1994; Hasbrouck, 1999; Hausman et al., 1992). Such market microstructure “frictions” importantly limit the efficiency of any realized volatility and related factor loadings constructed from the actual high-frequency data. A large and rapidly growing literature has sought to better characterize the general impact of these effects along with practical ways in which best to deal with the complications that arise in specific applications for different markets (see e.g., Andersen et al., 2000; Andreou and Ghysels, 2002; Bai et al., 2000; Barucci and Reno, 2002; Corsi et al., 2001; Epps, 1979; Lundin et al., 1999; Zhou, 1986; Zumbach et al., 2002). We will not pursue a thorough comparison of these alternative high-frequency-based realized volatility measures here.

[8] The estimate for $\beta_{t+1|t}$ defined by Eqs. (5)–(7) may alternatively be justified by the theory of quadratic variation along with an auxiliary assumption about the intraperiod continuous time dynamics underlying the discretely sample one-period returns in Eq. (1); see Zhang (2002) for further details.

Instead, following the analysis in Andersen et al. (2001b) of realized volatilities for individual stocks constructed from the TAQ database, we simply rely on interpolated 5-min equally spaced returns for all of our calculations.[9] Specifically, we record transaction prices for each of the individual stocks at each 5-min interval starting from 09:30 AM EST until 16:00 PM EST, by using the price of the trade at or immediately proceeding each 5-min mark.[10] With 79 5-min return observations per day, this results in more than 20,000 time series observations per stock per year over the 7-year sample. As in earlier studies in the literature, we only include firms with positive book values and at least 2 years of Compustat history.[11] On average, this leaves us with about 6400 stocks per year, for a total of close to *one billion* ($7 \times 20{,}000 \times 6{,}400$) 5-min return observations.

The 5-min continuously compounded returns on the market portfolio, MKT, are constructed as the logarithmic transform of the value-weighted percentage returns across all of the stocks. Following the approach in Fama and French (1993), the excess returns for the HML and SMB factor representing portfolios are defined by the return differential between the six portfolios grouped according to the individual stocks book-to-market ratios and their market capitalization. The returns on the 25 book-to-market and size-sorted test portfolios are calculated similarly.

Although the most liquid stocks in our sample generally trade multiple times within each 5-min interval, some of the less liquid stocks may not trade for up to an hour or longer. As such, the 5-min portfolio returns are clearly susceptible to nonsynchronous trading effects. It is well known that such effects may systematically bias the estimates of the factor loadings from traditional time series regressions like Eq. (7) above. Several different adjustment procedures have been proposed in the empirical asset pricing literature for dealing with this phenomenon over coarser daily and monthly frequencies (see, e.g., Andersen, 1991; Campbell et al., 1997, for a review of some of the classical studies in this literature). In particular, under the simplifying assumption that the trading intensity and the true latent process for the returns are independent, the one-factor CAPM adjustment procedure first proposed by Scholes and Williams (1977) works by accumulating additional leads and lags of the sample autocovariances between the returns and the market portfolio.[12] Adapting the Scholes–Williams procedure to a multi-factor framework, our empirical variation and covariation

[9] We also experimented with the use of coarser 10-, 15- and 30-min raw returns in calculating the realized volatilities and factor loadings. However, we found the results based on the 5-min returns coupled with the Scholes–Williams adjustment procedure discussed below to be preferable. Details of these results are available in Zhang (2002).

[10] Alternatively, we could have used the mid-point of the actual bid-ask quotes prevailing at each 5-min mark. However, the quotes are clearly not updated continuously, and will often be quite stale. Moreover, the quotes are also subject to strategic positioning by the dealer(s). Nonetheless, it is certainly possible that further experimentation along these lines may result in even better empirical results than the ones reported in Section 4.

[11] The book-to-market calculations are based on the Compustat and CRSP merged datasets. Stock splits and mergers are accounted for in the overnight returns, but dividends are excluded.

[12] The independence assumption is obviously a simplification. It would be interesting, but also computationally much more demanding, to explore more complicated adjustment procedures which break this link.

measures for the factors and the returns are based on the following generalizations of Eqs. (5) and (6),

$$\langle v,v\rangle_{t+1} = \sum_{j=-L,\ldots,L}\sum_{i=1,\ldots,1/h} f_{t+ih,h}f_{t+(1-j)h,h}{}' - 2Lh\left(\sum_{i=1,\ldots,1/h} f_{t+ih,h}\right) \times \left(\sum_{i=1,\ldots,1/h} f_{t+(i-j)h,h}\right)' \quad t=0,1,\ldots,T-1, \tag{8}$$

and,

$$\langle v,p\rangle_{t+1} = \sum_{j=-L,\ldots,L}\sum_{i=1,\ldots,1/h} f_{t+ih,h}r_{t+(1-j)h,h}{}' - 2Lh\left(\sum_{i=1,\ldots,1/h} f_{t+ih,h}\right) \times \left(\sum_{i=1,\ldots,1/h} r_{t+(i-j)h,h}\right)' \quad t=0,1,\ldots,T-1, \tag{9}$$

respectively, where L denotes the maximum lag length included in the adjustment, and $h=1/(22\times79)\approx5.75\times10^{-4}$ corresponds to our choice of intraday 5-min returns and a 1-month return horizon. The realized factor loadings are then simply obtained by direct substitution of $\langle\cdot,\cdot\rangle_{t+1}$ in place of $[\cdot,\cdot]_{t+1}$ in Eq. (7). Note that for $L=0$ the expressions in Eqs. (8) and (9) reduces to the standard expressions in Eqs. (5) and (6) underlying the conventional unadjusted OLS regression estimates.

In the empirical results reported on below we chose a lag window of $L=12$, or 1-h. This particular choice of L was motivated by visual inspection of the adjusted factor loadings, which showed little sensitivity to the choice of L beyond that lag. To illustrate, Fig. 1 plots the average factor loadings across the full sample associated with the MKT, SML and HML factor representing portfolios for the test portfolio termed 11 as a function of the lag length, L. Portfolio 11 (discussed in more detail in the next section) consists of the stocks in the smallest size and lowest book-to-market quintiles, and as such is among the test portfolios most prone to nonsynchronous trading effects. Nonetheless, it is evident from the figure that the average loadings for all three factors stabilize fairly quickly, and is very close to the average full-day lag 79 adjusted loadings for L around 12. A similar picture emerges for all of the other portfolios with the convergence to the full-day average occurring even faster (smaller values of L) in most other cases. We next turn to a discussion of the realized factor loadings for the three-factor Fama–French model and the 25 portfolios obtained by implementing this procedure.

3.2. Fama–French realized portfolio returns and factor loadings

The gains associated with the use of the high-frequency data in more accurately measuring and forecasting the realized factor loadings invariably depend upon the quality

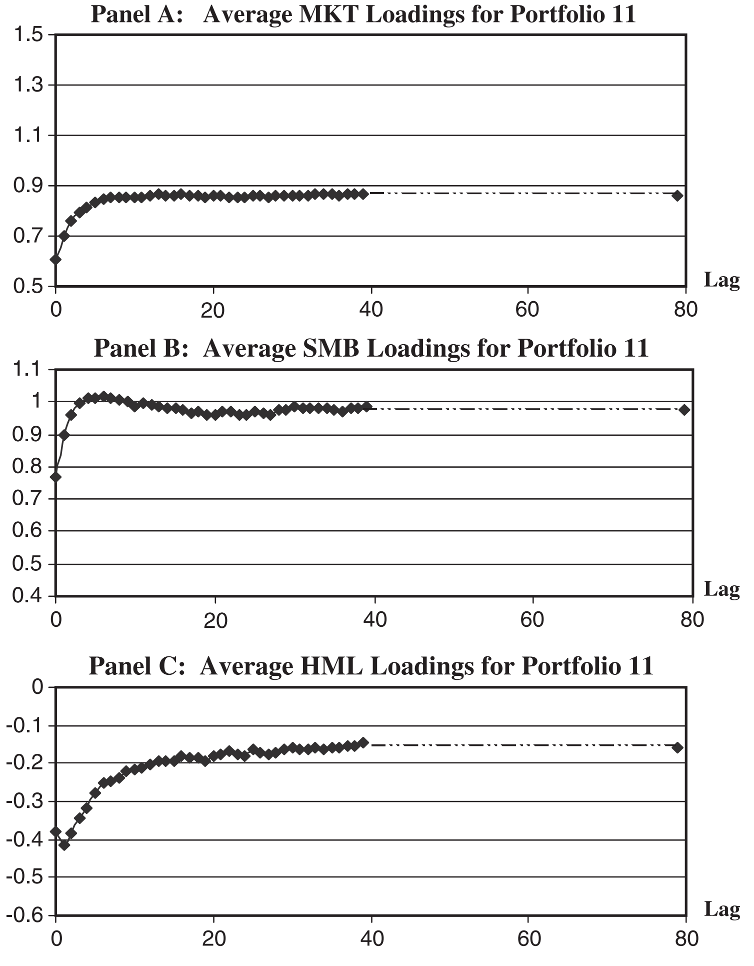

Fig. 1. The figure plots the average Scholes–Williams adjusted factor loadings for portfolio 11 as a function of the lag adjustment, $L = 0, 1, 2, \ldots, 40, 79$, as formally defined in Eqs. (7)–(9) in the main text. The sample consist of 5-min returns spanning the period from January 1993 through December 1999.

of the corresponding asset pricing model. The empirical results in the extant literature discussed above suggest that the asset pricing anomalies related to size and book-to-market largely disappear within the context of the three-factor Fama–French model. As such, this particular model and the 25 test portfolios sorted on the basis of firm size and book-to-market ratios provide a natural setting for evaluating the improvements afforded by the high-frequency-based factor loadings.

Following standard convention in the literature, we number the 25 portfolios from 11 to 55, with the first and second index referring to the size and book-to-market quintiles, respectively. The average monthly percentage returns and realized volatilities, or monthly standard deviations, over the 7-year (1993–1999) sample period are given in Table 1. All of the average returns are positive, and except for the highest book-to-market portfolios,

Table 1
Monthly portfolio returns

	Low	2	3	4	High
Mean returns					
Small	0.765	1.654	1.527	1.826	1.853
2	1.308	1.202	1.223	1.082	1.208
3	1.243	1.171	1.169	1.005	1.333
4	1.564	1.010	1.030	0.988	0.890
Big	1.630	1.284	1.323	1.287	1.075
Realized volatility					
Small	4.407	3.779	3.104	2.750	2.773
2	5.152	3.758	2.951	2.769	3.144
3	5.201	3.674	3.167	3.010	3.170
4	5.248	3.565	3.294	2.933	2.824
Big	4.179	3.794	3.841	3.697	3.484

The table reports the average monthly percentage returns and realized volatility for the 25 size and book-to-market sorted portfolios over the January 1993 through December 1999 sample period. The realized monthly volatilities are reported in standard deviation format, and constructed from the summation of the squared 5-min returns and the 1-day lead-lag autocovariances.

slightly higher than the returns over the longer 1963–1991 time period analyzed in Fama and French (1993). In contrast to the earlier results, the cross sectionally averages no longer show a clear-cut pattern in regards to the size and book-to-market characteristics. As noted by Cochrane (2001), among others, small company stocks with low book-to-market ratios did relatively poorly during most of the 1990s, in effect reversing the size premium. Also, in a reversal of the value premium, growth stocks (low book-to-market value) in the largest size quantile did better than value stocks (high book-to-market value) of the same size. Meanwhile, the average realized volatilities over the recent 7-year period are generally in line with the long-run historical sample standard deviations, with the portfolios in the lowest book-to-market quintile exhibiting the highest average return volatility.

The reversal of the size and value premiums is also evident from the average monthly returns on the HML and SMB zero investment portfolios. Whereas both exhibit positive average returns over the longer historical sample, the average monthly percentage returns over the 7-year time period analyzed here are -0.107 and -0.122, respectively. Of course, this does not invalidate the explanatory power of the three-factor Fama–French model. Indeed, as noted below, the model still yields an average R^2 in excess of 50% for explaining the cross-sectional differences in the monthly returns across the 25 portfolios—more than twice the average R^2 for the standard one-factor CAPM over the same time period. Also, the average R^2 for the 25 time series regressions of the portfolio returns on the three factors exceeds 85%. These results are directly in line with the results reported in the existing literature based on longer historical samples.

Turning to Table 2, Panel A reports the monthly sample means for each of the 25 realized portfolio loadings. The average loadings for the SMB factor systematically decrease with the size of the portfolio, while the average realized loadings for the HML factor increase with the book-to-market ratio. No apparent pattern is evident in the loadings association with the return on the market portfolio, MKT.

Table 2
Realized factor loadings

Panel A: sample means	Low	2	3	4	High
MKT					
Small	0.869	1.099	0.990	1.063	1.099
2	1.093	0.973	0.932	0.875	1.020
3	1.074	0.936	0.933	0.887	1.039
4	1.115	0.874	0.894	0.853	0.816
Big	0.959	0.956	1.027	1.058	1.000
SMB					
Small	1.001	1.133	1.000	1.042	1.072
2	1.065	0.923	0.848	0.801	0.899
3	0.834	0.743	0.716	0.656	0.715
4	0.520	0.362	0.385	0.332	0.213
Big	− 0.202	− 0.192	− 0.209	− 0.140	− 0.204
HML					
Small	− 0.187	0.243	0.378	0.600	0.727
2	− 0.388	0.047	0.358	0.424	0.608
3	− 0.439	0.082	0.316	0.425	0.656
4	− 0.471	0.075	0.293	0.412	0.502
Big	− 0.439	0.024	0.367	0.585	0.782
Panel B: sample standard deviations					
MKT					
Small	0.181	0.105	0.102	0.120	0.125
2	0.109	0.090	0.080	0.075	0.095
3	0.135	0.114	0.102	0.136	0.125
4	0.115	0.107	0.114	0.116	0.105
Big	0.068	0.103	0.134	0.113	0.132
SMB					
Small	0.278	0.177	0.140	0.179	0.188
2	0.201	0.162	0.143	0.126	0.152
3	0.247	0.205	0.176	0.201	0.225
4	0.253	0.173	0.177	0.175	0.178
Big	0.113	0.137	0.165	0.149	0.219
HML					
Small	0.288	0.233	0.231	0.198	0.217
2	0.192	0.209	0.173	0.164	0.205
3	0.335	0.218	0.171	0.244	0.349
4	0.281	0.239	0.246	0.250	0.289
Big	0.182	0.247	0.269	0.272	0.279

(continued on next page)

Table 2 (*continued*)

Panel C: first-order autocorrelations

	Low	2	3	4	High
MKT					
Small	0.294	0.065	0.317	0.477	0.302
2	0.199	0.172	0.244	0.453	0.434
3	0.360	0.492	0.617	0.512	0.208
4	0.310	0.237	0.368	0.360	0.287
Big	0.359	0.553	0.301	0.329	0.320
SMB					
Small	0.312	0.101	0.109	0.170	0.026
2	0.345	0.115	0.365	0.288	0.130
3	0.654	0.541	0.420	0.302	0.115
4	0.300	0.217	0.200	0.227	0.279
Big	0.409	0.136	0.125	0.167	0.429
HML					
Small	0.306	0.134	0.252	0.232	0.128
2	0.379	0.367	0.203	0.352	0.172
3	0.470	0.298	0.166	0.378	0.086
4	0.279	0.551	0.520	0.483	0.222
Big	0.567	0.691	0.255	0.597	0.246

Panel A reports the average realized factor loadings for the three-factor Fama–French model for each of the 25 size and book-to-market sorted portfolios over the January 1993 through December 1999 sample period. Panel B gives the corresponding sample standard deviations, while Panel C reports the first-order sample autocorrelations. The realized factor loadings are constructed from high-frequency 5-min returns as discussed in the main text of the paper.

The monthly sample standard deviations reported in Panel B suggest that the realized factor loadings also vary importantly through time, with the loadings for the largest low book-to-market portfolios generally being the least variable. The existence of important temporal variation in the factor loadings is further underscored by the first-order sample autocorrelations reported in Panel C.[13] In spite of the limited 84-month sample, the vast majority of the autocorrelations are statistically significant when judged by the usual asymptotic Bartlett standard error band. Similarly, Ljung–Box portmanteau tests for up to 10th order serial correlation (available upon request) are significant for the vast majority of the portfolios. The positive own serial correlation is also manifest in the time series plots for the realized factor loadings for portfolio 55 given in Fig. 2. Even though the inherent measurement errors result in somewhat erratic month-to-month realized factor loadings, an underlying systematic

[13] As discussed in Section 3.1, the 5-min realized factor loadings are invariably subject to measurement errors. This will inflate the reported standard deviations in Panel B, and result in a downward bias in the sample autocorrelations in Panel C. It would be interesting, but beyond the scope of the present paper, to formally assess the magnitude of the measurement errors in the present context (Bai et al., 2000; Barndorff-Nielson and Shephard, 2002a,b; Meddahi, 2002, all provide important theoretical results for the case of scalar realized volatility measures).

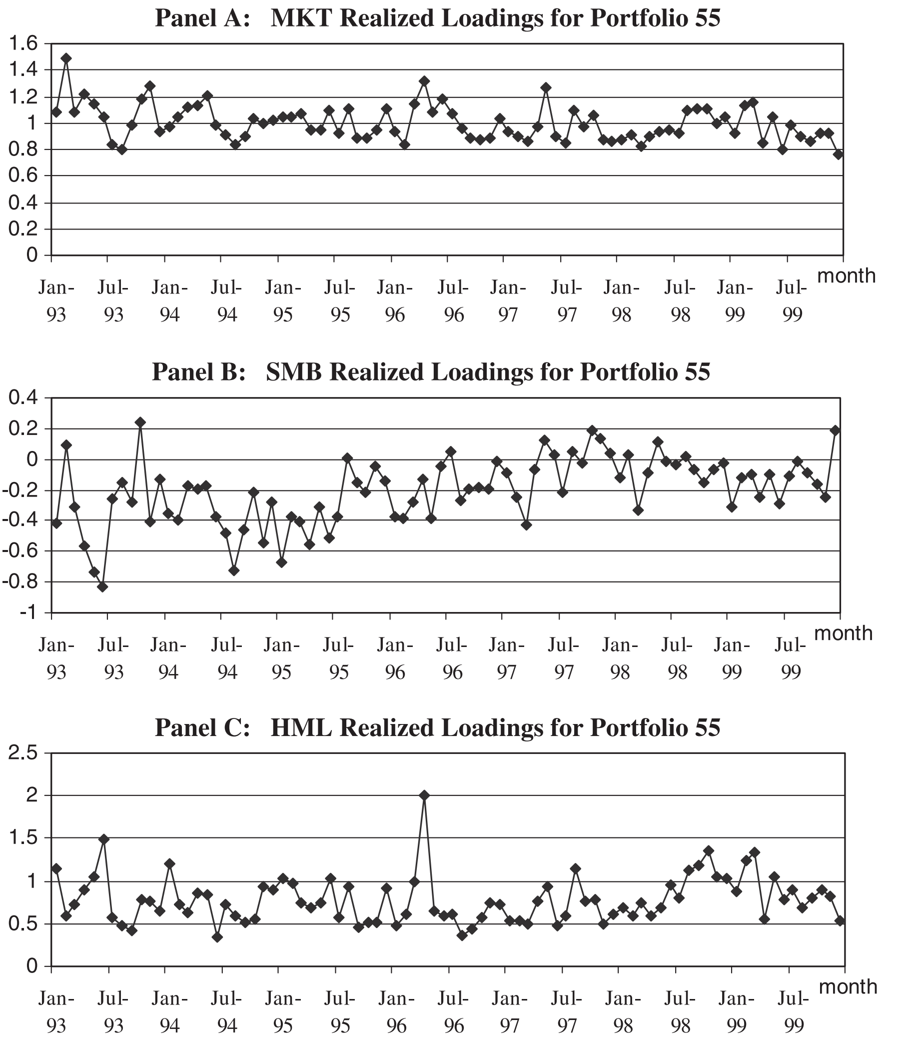

Fig. 2. The figure plots the realized Scholes–Williams adjusted factor loadings (based on the inclusion of 12 leads and lags) for portfolio 55. The sample spans the 84-month time period from January 1993 through December 1999.

temporal pattern is evident for all three risk factors. Similar results hold true for the other portfolios.

The realized factor loadings are, of course, only observable ex post. The results in the next section demonstrate that a simple AR(1) time series model for the high-frequency factor loadings effectively smoothes out the measurement error, and gives rise to more accurate forecasts and statistically significant lower pricing errors when compared to the conventional monthly rolling regression-based estimates traditionally employed in the literature.

4. High-frequency factor loadings and asset pricing

The discussion in the previous section illustrated how high-frequency data may be used in an actual empirical setting for more accurately measuring systematic factor risk exposures. This section goes one step further and asks the question: do the high-frequency-based measures give rise to more accurate predictions of the future factor loadings and statistically significant improvements in terms of asset pricing predictions and corresponding better economic decisions? We begin by a discussion of our four different return-based forecast procedures for the factor loadings—two based on monthly return observations only, two directly utilizing the high-frequency measures.

4.1. Predicted factor loadings

Our first procedure takes the factor loadings to be constant and equal to the conventional regression estimates based on 5 years of monthly observations at the beginning of the out-of-sample forecast comparison period.[14] We will refer to these constant factor loadings by the acronym CON. We also experimented with the loadings estimates based on the monthly return observations over the full sample. Of course, from a forecasting perspective these are formally in-sample estimates, and would not actually be available until the end of the sample period. Nonetheless, very similar results to the ones reported below for the CON procedure were obtained using these estimates. Details of these results are available upon request.

The 5-year rolling regression-based estimates advocated by Fama and MacBeth (1973), and adopted in numerous subsequent studies, allows the betas to change every month as new return observations replace the return observations at the start of the sample. In the discussion below, we will refer to these conventional rolling regression-based estimates by the acronym RR (for a more formal analysis of this procedure, see also Foster and Nelson, 1996). As previously noted, on utilizing this procedure along with the actual realized factor returns, the time series average of the 84 monthly cross-sectional R^2's for explaining the variation in the returns across the 25 portfolios equals 0.529, whereas the average R^2 from the 25 time series regressions equals 0.856.

As a simple benchmark, our first high-frequency-based procedure takes the forecast for the next month to be equal to the current month's realized loadings. These forecasts would, of course, be optimal in a mean-square-error sense if the time series of realized factor loadings followed a Random Walk (RW), or martingale, model. We consequently refer to this procedure as the HF-RW-based forecasts.

Our second, and final, high-frequency procedure is based on the one-step-ahead forecasts from a set of AR(1) models for each of the different loadings and portfolios. The estimates of the models are based on a rolling 24-month sample, leaving us with an

[14] In order to implement these regressions, and the rolling regressions discussed below, we augmented the January 1993 through December 1999 high-frequency data, with 5 years of monthly returns on the 25 portfolios and three factors from January 1988 through December 1992. These portfolio and factor returns are available from Kenneth French's website, http://web.mit.edu/kfrench/www/data_library.html.

Table 3
Factor loading forecasts

	CON	RR	HF-RW	HF-AR
MKT				
Bias	− 0.071	− 0.039	0.001	0.007
MSE	0.039	0.026	0.017	0.012
MAE	0.157	0.125	0.100	0.086
SMB				
Bias	0.065	0.009	0.003	0.025
MSE	0.066	0.046	0.049	0.033
MAE	0.201	0.165	0.170	0.140
HML				
Bias	0.095	0.059	0.004	0.035
MSE	0.072	0.067	0.061	0.045
MAE	0.211	0.202	0.186	0.161

The table reports the average forecast errors for forecasting the monthly realized factor loadings across the 25 portfolios and the 5-year out-of-sample period from January 1995 through December 1999. The predictions labeled CON takes the factor loadings to be constant and equal to the conventional 5-year monthly regression-based estimates at the beginning of the sample. RR refers to the 5-year rolling regression estimates. HF-RW assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The HF-AR predictions are based on a set of rolling AR(1) models estimated over the previous 24 months. The rows labeled Bias, MSE, and MAE give the average errors, the average squared errors, and the average absolute errors, respectively.

out-of-sample forecast comparison period covering the 60 months from January 1995 through December 1999. Consistent with the significant autocorrelations reported in Table 3, the vast majority of the parameter estimates for the $3 \times 25 \times 60 = 4500$ different AR(1) models are statistically significant at standard levels.[15] In addition to these rolling AR(1) models, we also experimented with the forecasts based on a set of ARMA(1,1) models estimated over the full in-sample period.[16] The basic findings, available upon request, were very similar to the ones reported below for the AR(1) models. Meanwhile, more complicated dynamic time series models, possibly allowing for asymmetries in the relationship between the past returns and the future factor loadings, may give rise to even better predictions (Ball and Kothari, 1989; Braun et al., 1995; Cho and Engle, 1999, have previously documented statistically significant temporal variation and important asymmetries in CAPM betas, while Zhang, 2002, provide complementary evidence for asymmetric market risk exposure based on the same type of high-frequency measures utilized here). To keep the comparisons simple, we purposely do not pursue any of these

[15] In order to account for the small sample biases in the estimated AR(1) coefficients, we follow the classical procedure of Orcutt and Winokur (1969) in basing our model predictions on the adjusted coefficients $(N\varphi + 1)(N - 3)^{-1}$, where $N = 24$, and φ denotes the OLS estimate. Details of the estimates are available upon request.

[16] The ARMA(1,1) formulation is consistent with the assumption that the true latent loadings follow an AR(1) model, but that the (observed) realized loadings are equal to the true loadings plus an independent measurement error.

more complicated model specifications here. Of course, the relatively short 7-year time span of high-frequency data also limits the scope of such investigations.

Turning to the results, Table 3 reports the average bias (Bias), average squared error (MSE), and average absolute error (MAE), using each of the four procedures for forecasting the monthly realized loadings. The errors are averaged across the 25 size and book-to-market sorted portfolios and the 60-month out-of-sample period. The loadings associated with the market return, MKT, are generally the most accurate. Not surprisingly, the constant factor loadings (CON) result in the largest average Biases, MSEs and MAEs of all the procedures. The MSEs and MAEs for the RR and HF-RW procedures are fairly close. The one-step-ahead AR(1) forecasts systematically result in the lowest MSEs and MAEs for all of the factor loadings. Intuitively, the HF-AR-based procedure more swiftly adjusts to changes in the underlying market conditions.

This is further illustrated in Fig. 3, which plots the RR- and HF-AR-based loadings for each of the three factors with respect to portfolio 55. As seen from the figure, the sensitivity to the MKT factor measured by the RR procedure starts out at a relatively high value of 1.3, but eventually adjusts to a value around unity, consistent with the predictions by the HF-AR procedure. Similarly, the two SML loadings start out at very different values, only to converge at the end of the sample. Meanwhile, the ordering of the RR- and HF-AR-based loadings for the HML factor is reversed halfway through the sample. These particular results are, of course, specific to portfolio 55. However, equally pronounced (albeit qualitatively different) orderings obtain for the other portfolios.

It would be interesting to relate the documented temporal dependencies in the high-frequency return-based factor loadings to specific macroeconomic influences. However, as previously noted, the scope of such investigations is invariably limited by the short time span of the data. Thus, following Ferson and Harvey (1999), we simply focus on the same five explanatory variables found to play an important role in their construction of scaled factor loadings over a much longer 30-year monthly sample, namely, the spread between 1- and 3-month T-bill returns (hb3), the S&P500 dividend yield (div), Moody's Aaa and Baa corporate bond yield spread (junk), the spread between 1- and 10-year T-bonds (term), and the return on a 30-day T-bill (Tbill). Consistent with the earlier empirical results in Ferson and Harvey (1999), these five instrumental variables explain between 5% and 10% of the monthly return variability in the 25 test portfolios for the 1993–1999 sample. More interesting from the present perspective, however, regressing the high-frequency-based realized monthly loadings on the five instrumental variables result in their joint significance for 14, 11 and 6 of the 25 regressions for the MKT, SMB, and HML factor loadings, respectively. Also, the average R^2's for the three different factor loadings equal 0.124, 0.134, and 0.206, respectively. These R^2's are comparable to the average values from the AR(1) models of 0.135, 0.091, and 0.140, respectively, and somewhat lower than the 0.215, 0.161, and 0.232, respectively, obtained by including both the own lagged loadings and the instrumental variables in the regressions.[17] As such, these results suggest that the information in the five instrumental variables (in part) complements the information contained in the own lagged loadings. We will not pursue this approach any further here,

[17] Further details concerning these results are available in Zhang (2002).

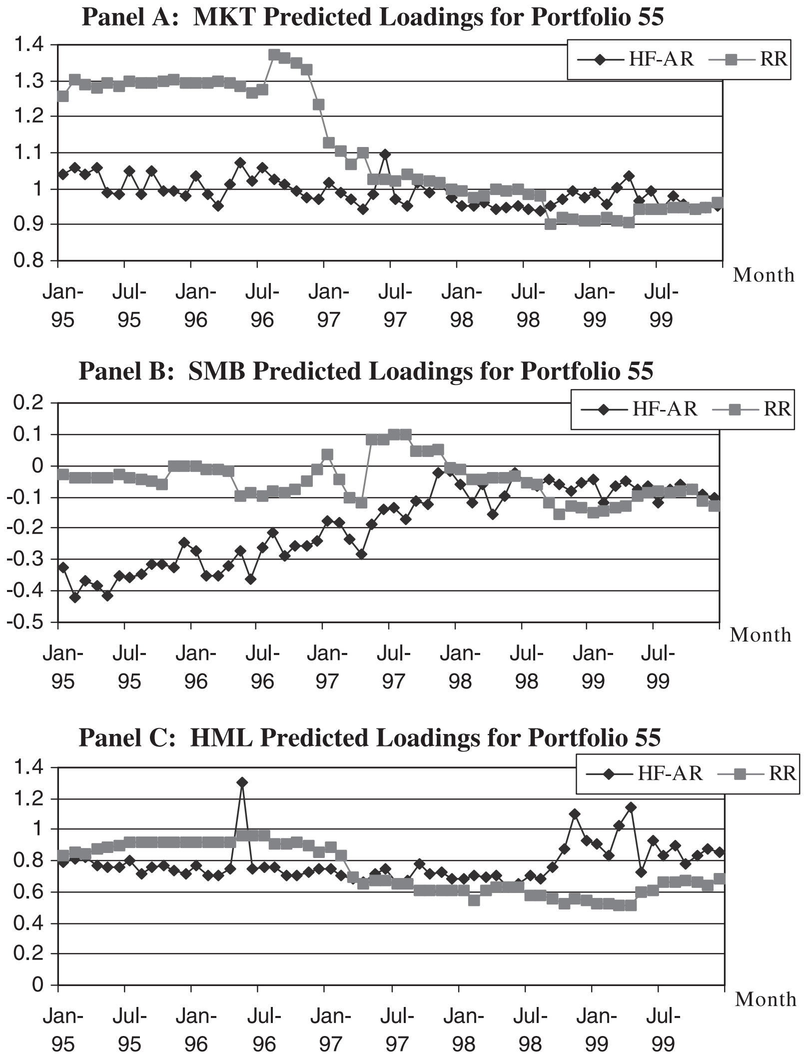

Fig. 3. The figure plots the RR- and HF-AR-based predictions, as described in the main text, for the factor loadings for portfolio 55. The forecasting period consists of the 60 months from January 1995 through December 1999.

however. Instead, we next turn to a discussion of the asset pricing improvements afforded by the simple purely return-based high-frequency factor loadings.

4.2. Asset pricing

The prediction of asset prices is notoriously difficult, and any temporal variation in the factor loadings are invariable of second order importance relative to the temporal variation in the associated factor risk premiums (see, for instance, Cochrane, 2001). Hence, in order to focus directly on the potential benefits of using the high-frequency data in measuring and modeling the realized factor loadings, we rely on the actual realized excess factor returns when calculating the asset prices.[18]

The three-factor Fama–French model has been the subject of numerous empirical studies and, as discussed above, has been found to perform very well in pricing the 25 size and book-to-market portfolios analyzed here. Indeed, for none of the different factor loading forecast procedures is the hypothesis of a zero "alpha" in the three-factor model rejected by a simple t-test. Specifically, let the forecast error for the returns on portfolio i for month t based on the loadings from procedure m, be denoted by $e_{i,t}(m)$. The four t-statistics for testing the hypotheses that the average pricing errors across the 25 portfolios and the 60-month out-of-sample period are equal to zero (i.e., $\sum_i \sum_t e_{i,t}(m) = 0$, then take the values -0.144, -0.092, 0.002, and -0.006 for the CON-, RR-, HF-RW- and HF-AR-based loadings, respectively. The failure to reject the three-factor model for any of the different procedures by this test over such a limited 5-year period is, of course, not surprising.

Meanwhile, the superiority of the high-frequency-based loadings is underscored by the results in Table 4, and the average absolute pricing errors, $|e_{i,t}(m)|$, reported in the first panel. Of course, the most appropriate statistical loss function for assessing the superiority of the competing forecasts will ultimately depend upon the application at hand (see e.g., Christoffersen and Diebold, 1997 for a discussion of different loss functions in the comparison of nonlinear forecasting models). Hence, following Andersen et al. (1999), Table 4 also reports the heteroskedasticity adjusted average absolute standardized errors, $|e_{i,t}(m)/\sigma_{i,t}|$, and the pseudo R^2 defined by $1-(e_{i,t}(m)/\sigma_{i,t})^2$, where and $\sigma_{i,t}$ refers to the realized volatility for portfolio i for month t constructed by the summation of the 5-min squared returns within the month augmented by the 1-day lead-lag autocovariances to account for the nonsynchronous trading effects. In order to conserve space, we only report the average loss across the 25 portfolios. From the results, reported in the first rows in each of the three panels, it is evident that the HF-AR-based forecasts systematically yield the lowest absolute errors and the highest pseudo R^2's among the four different procedures.[19]

To formally test for the statistical significance of these apparent differences, Table 4 complements the average loss with the values of the three relatively simple pair-wise forecast comparison tests discussed by Diebold and Mariano (1995). In particular, let $d_{i,t}(m,n)$ denote the loss differential between the forecasts based on method m and n for portfolio i and month t. The $Z1_{m,n}$ statistic then refers to the standard t-test for the hypothesis that the mean differential is equal to zero. Alternatively, the null hypothesis of equal forecast accuracy, may also be tested by the nonparametric sign-test,

$$Z2_{m,n} = \left[\sum_i \sum_t I_+(d_{i,t}(m,n)) - 0.5T\right][0.25T]^{1/2}, \tag{10}$$

[18] Since the errors associated with the forecasts of the factor loadings and the risk premiums are likely to be correlated, a rigorous true out-of-sample forecast comparison would also necessitate the joint modeling of the loadings and the factor risk premiums.

[19] Although the improvements may seem numerically small, the increase in the R^2 for the HF-AR procedure is actually somewhat higher than the improvements reported in the aforementioned study by Ferson and Harvey (1999) based on explicitly modeling the temporal variation in the monthly factor loadings using exogenous explanatory variables.

where $I_+(\cdot)$ denotes the indicator function and $T = 60 \times 25$ refers to the sample size. Lastly, the related studentized version of the Wilcoxon's signed-rank test is formally defined by,

$$Z3_{m,n} = \left[\sum_i \sum_t I_+(d_{i,t}(m,n))\text{rank}(\,|\,d_{i,t}(m,n)\,|\, - 0.25T(T+1)\right] \times [T(T+1)(2T+1)/24]^{-1/2}. \tag{11}$$

Under the null hypothesis that models m and n have the same predictive ability, all three test statistics should be asymptotically standard normally distributed. The values reported in the three lower panels in the table soundly reject the hypothesis of no superior predictive ability in favor of the high-frequency-based HF-AR procedure.[20]

In order to further explore the performance of the HF-AR procedure, Table 5 reports the average pseudo R^2's for each of the 25 portfolios. Not surprisingly, the R^2's are generally higher for the big stock portfolios, with the largest stock and lowest book-to-market portfolio having the highest overall R^2 of 0.949. Comparing the results to the R^2's associated with the conventional monthly rolling regression-based estimates, it follows from the second panel in the table, that the HF-AR procedure does better in an absolute sense for 17 out of the 25 portfolios. Interestingly however, for none of the portfolios where the RR procedure does better, is the difference statistically significant at the usual 5% level, whereas the HF-AR-based loadings result in statistically significant improvements for four of the portfolios.[21] Thus, in spite of the relatively short 5-year out-of-sample forecast comparison period, our various statistically tests systematically favor the asset pricing predictions based on the high-frequency factor loadings.

4.3. Economic significance of high-frequency loadings

To highlight the economic value of the high-frequency-based loadings and corresponding asset pricing predictions, this section compares the ex post performance of the mean-variance efficient portfolios constructed from the 25 size and book-to-market sorted portfolios and the same four forecasting procedures discussed above. In order to focus the comparisons directly on the impact of the forecasts for the factor loadings, we assume that the covariance matrix for the 25 test portfolios underlying the mean-variance optimizations is the same across the different procedures, and equal to the realized covariance matrix for the 24-month pre-sample period from January 1993 through December 1994.[22] Also, we explicitly exclude the risk-free asset from the portfolio optimizations.

[20] The results reported in Table 4 are based on the 60-month out-of-sample forecast comparison period and the simple AR(1) model. Qualitatively very similar results obtain for the full 84 month sample and the more complicated ARMA(1,1) model. For instance, the $Z1$, $Z2$, and $Z3$ statistics for comparing the pseudo R^2 for the RR and HF-ARMA(1,1) procedure based on the full in-sample parameter estimates equal 3.304, 2.138 and 2.491, respectively.

[21] The results for the two nonparametric $Z2$ and $Z3$ tests are very similar. For instance, 4 of the 25 $Z2_{\text{RR,HF-AR}}$ test statistics exceed 1.96 while none are less than -1.96.

[22] Comparable results for other covariance matrix estimators are available in Zhang (2002). The results based on a constant covariance matrix provides a lower bound on the benefits from using the high-frequency data, as further improvements are likely to obtained by also modeling the temporal variation in the realized covariance matrix (e.g., Fleming et al., in press). However, doing so in the context of a 25-dimensional covariance matrix presents a number of difficult empirical challenges.

The first panel in Table 6 reports the resulting average monthly realized returns over the ex post January 1995 through December 1999 sample period for the optimized portfolios with the monthly expected standard deviations indicated in the first column. For all of the

Table 4
Average pricing errors

	CON	RR	HF-RW	HF-AR
Panel A: absolute errors				
	1.360	1.259	1.307	1.243
$Z1_{\mathrm{CON,j}}$		5.689	1.778	4.852
$Z2_{\mathrm{CON,j}}$		4.803	4.028	6.094
$Z3_{\mathrm{CON,j}}$		6.367	3.457	6.362
$Z1_{\mathrm{RR,j}}$			−1.885	0.875
$Z2_{\mathrm{RR,j}}$			0.465	2.685
$Z3_{\mathrm{RR,j}}$			−0.098	2.405
$Z1_{\mathrm{HF\text{-}RW,j}}$				3.682
$Z2_{\mathrm{HF\text{-}RW,j}}$				2.272
$Z3_{\mathrm{HF\text{-}RW,j}}$				3.042
Panel B: absolute standardized errors				
	0.433	0.394	0.396	0.377
$Z1_{\mathrm{CON,j}}$		7.407	3.974	7.030
$Z2_{\mathrm{CON,j}}$		4.803	4.028	6.094
$Z3_{\mathrm{CON,j}}$		6.777	4.047	6.907
$Z1_{\mathrm{RR,j}}$			−0.241	2.851
$Z2_{\mathrm{RR,j}}$			0.465	2.685
$Z3_{\mathrm{RR,j}}$			0.223	3.144
$Z1_{\mathrm{HF\text{-}RW,j}}$				3.401
$Z2_{\mathrm{HF\text{-}RW,j}}$				2.272
$Z3_{\mathrm{HF\text{-}RW,j}}$				2.897
Panel C: pseudo R^2				
	0.646	0.715	0.711	0.743
$Z1_{\mathrm{CON,j}}$		7.477	3.834	7.044
$Z2_{\mathrm{CON,j}}$		6.094	4.028	6.094
$Z3_{\mathrm{CON,j}}$		7.309	4.269	7.448
$Z1_{\mathrm{RR,j}}$			−0.349	3.183
$Z2_{\mathrm{RR,j}}$			0.465	2.685
$Z3_{\mathrm{RR,j}}$			−0.155	3.322
$Z1_{\mathrm{HF\text{-}RW,j}}$				3.410
$Z2_{\mathrm{HF\text{-}RW,j}}$				2.272
$Z3_{\mathrm{HF\text{-}RW,j}}$				3.162

Table 5
Pseudo R^2

	Low	2	3	4	High
Sample means					
Small	0.550	0.619	0.626	0.586	0.480
2	0.836	0.888	0.815	0.753	0.710
3	0.833	0.772	0.654	0.664	0.713
4	0.846	0.824	0.736	0.743	0.689
Big	0.949	0.890	0.860	0.874	0.656
$Z1_{RR,HF\text{-}AR}$					
Small	− 0.843	− 1.378	1.432	2.724	1.567
2	− 1.625	0.890	1.440	− 0.981	1.966
3	1.278	2.043	− 1.675	1.326	0.601
4	− 0.267	1.025	− 0.208	− 0.437	2.932
Big	1.900	1.265	1.876	0.594	1.227

The first panel reports the pseudo R^2 for each of the 25 portfolios defined by one minus the squared standardized error for the HF-AR predictions averaged across the 5-year out-of-sample period from January 1995 through December 1999. The HF-AR predictions are based on a set of rolling AR(1) models estimated over the previous 24 months. The second panel reports the *t*-statistic for testing identical mean R^2's between the 5-year rolling regression estimates, RR-, and HF-AR based predictions. Under the null hypothesis of no superior predictive ability by either of the models, the test statistics should be standard normally distributed. See the main text for further details.

monthly standard deviations ranging from 0.02 to 0.06, or 0.07 to 0.21 at an annual level, the ex post returns are the highest for the HF-AR-based factor loadings. With the exception of the low-risk portfolio with a monthly expected standard deviation of 0.02, all of the Sharpe ratios for the monthly ex post returns and standard deviations, reported in the second panel in Table 6, also favor the HF-AR procedure by a fairly large margin. The very high ex post Sharpe ratios reflect the unprecedented bull-run during the 5-year out-of-sample period. Although it would be unrealistic to expect these same high values to obtain over other time periods, the ordering among the different procedures are nonetheless suggestive.

Notes to Table 4:
Panel A reports the average absolute errors. The entries in the first row have been multiplied by $\times 10^2$. Panel B gives the average absolute standardized errors obtained by standardizing each error by realized volatility for that particular portfolio and month. Panel C reports the average pseudo R^2 defined by one minus the squared standardized error. All of the averages are calculated over the 25 portfolios and the 5-year out-of-sample period from January 1995 through December 1999. The procedure labeled CON takes the factor loadings to be constant and equal to the conventional monthly regression estimates at the beginning of the sample. RR refers to the 5-year rolling regression estimates. HF-RW assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The HF-AR predictions are based on a set of rolling AR(1) models estimated over the previous 24 months. $Z1_{i,j}$ refers to the *t*-test for a difference in the means across the $25 \times 60 = 1500$ monthly predictions for models *i* and *j*. $Z2_{i,j}$ gives the sign test while $Z3_{i,j}$ refers to the Wilcoxon's signed rank test for the same hypothesis. Under the null hypothesis of no superior predictive ability by either of the two models, all of the test statistics should be standard normally distributed. See the main text for further details.

Table 6
Mean-variance efficient portfolios

SD	CON	RR	HF-RW	HF-AR
Ex post returns				
0.02	0.037	0.042	0.040	0.043
0.03	0.063	0.073	0.069	0.075
0.04	0.085	0.099	0.094	0.102
0.05	0.105	0.124	0.117	0.128
0.06	0.125	0.148	0.140	0.153
Sharpe ratios				
0.02	0.925	1.095	0.984	1.071
0.03	1.081	1.323	1.241	1.431
0.04	1.091	1.337	1.306	1.546
0.05	1.081	1.323	1.326	1.594
0.06	1.069	1.306	1.332	1.616

The table reports the average monthly ex post percentage returns and corresponding Sharpe ratios for the mean-variance efficient portfolios constructed from the 25 size and book-to-market sorted test portfolios over the 5-year out-of-sample period from January 1995 through December 1999. The expected monthly standard deviations reported in the first column are based on the pre-sample realized variance–covariance matrix constructed from 5-min returns from January 1993 through December 1994. The expected return calculations labeled CON takes the factor loadings to be constant and equal to the conventional 5-year monthly regression-based estimates at the beginning of the sample. RR refers to the 5-year rolling regression estimates. HF-RW assumes that the high-frequency-based realized factor loadings follow a random walk, or martingale, model. The HF-AR predictions are based on a set of rolling AR(1) models estimated over the previous 24 months.

Our final illustration of the economic significance of the high-frequency-based loadings is provided in Table 7. Following the recent analysis in Fleming et al. (2001, 2003), the table reports the monthly fees that an investor with a quadratic utility function and constant relative risk aversion equal to γ would be willing to pay to switch from the traditional RR 5-year monthly rolling regression-based forecasts to the high-frequency-based HF-AR forecasts for the 25 portfolio returns. The fees in the

Table 7
Willingness to pay

SD	$\gamma=1$	$\gamma=10$
0.02	0.10	0.09
0.03	0.22	0.23
0.04	0.34	0.37
0.05	0.46	0.52
0.06	0.60	0.69

The table reports the monthly fee (in percentage points) that an investor with a quadratic utility function and constant relative risk aversion of γ would be willing to pay to switch from the 5-year rolling regression-based return forecasts (RR) to the high-frequency-based AR(1) forecasts for the realized factor loadings (HF-AR) for the 25 size and book-to-market sorted test portfolios. The fees are based on the monthly returns for the 5-year out-of-sample period from January 1995 through December 1999, with the monthly expected standard deviations for both methods determined by the pre-sample realized covariance matrix from January 1993 through December 1994.

table are expressed in monthly percentage points and determined empirically as the value which equalizes the ex post utility for the two different procedures when evaluated on the basis of the monthly returns from January 1995 through December 1999.[23] To facilitate comparisons, we again fix the expected covariance matrix for both procedures at the realized covariance matrix for the January 1993 through December 1994 pre-sample period. The numbers are telling. For instance, fixing the expected standard deviation at 0.04, or 14% at an annual level, an investor with $\gamma = 10$ would be willing to pay 0.37, or almost 4.5% per year, to switch from the RR to the HF-AR monthly return-based forecasts.

5. Conclusion

The implementation of multi-factor pricing models has traditionally been plagued by an errors-in-variables problem stemming from the use of estimated factor loadings. Focusing on the three-factor Fama–French model and a set of size and book-to-market sorted portfolios, this paper demonstrates how high-frequency intraday transactions prices for the different portfolios and the underlying factor returns may be used in more effectively measuring and modeling the realized factor loadings.

Several interesting directions for future research remain. First, in order to facilitate comparisons with the existing academic finance literature, we have purposely focused our empirical analysis on a 1-month return horizon and the Fama–French test portfolios. It would be interesting to explore the robustness of the results with respect to other portfolios and return horizons. In this regard, the use of daily or weekly realized factor loadings may result in important improvements over shorter return horizons of immediate interest to finance practitioners concerned with the practical implementation of competing hedging and/or trading strategies. Similarly, it would also be interesting to see how the results carry over to other markets. Second, the forecasting results reported here were based on very simple time series models. Presumably, even better results may be obtained through the implementation of more refined models for the temporal variation in the individual realized factor loadings, possibly allowing for asymmetries in the relationship between the loadings and the past returns. Along these lines, it would also be interesting to further investigate how the temporal variation in the factor loadings relates to the underlying macroeconomic influences. Of course, the lack of reliable high-frequency data over long time spans invariably limits the scope of such investigations. Third, the development of new and improved methods for dealing with the market microstructure complications inherent in the data are likely to further enhance the benefits of all high-frequency-based risk measures. We leave all of these issues for future research.

[23] The fee, Δ, is formally determined by solving $\sum_{t=1,\ldots,T}\{(r_{t,\mathrm{HF-AR}} - \Delta) - [\gamma/2(1+\gamma)](r_{t,\mathrm{HF-AR}} - \Delta)^2\} = \sum_{t=1,\ldots,T}\{r_{t,\mathrm{RR}} - [\gamma/2(1+\gamma)]r^2_{t,\mathrm{RR}}\}$, where $r_{t,\mathrm{HF\text{-}AR}}$ and $r_{t,\mathrm{RR}}$ refer to the ex post monthly returns for the mean-variance efficient portfolio with the same benchmark expected standard deviations and expected returns determined by the HF-AR and RR loading forecasts, respectively.

Acknowledgements

This work was supported by a grant from the National Science Foundation to the NBER. The main idea behind the paper has been importantly influenced by conversations and related joint work with Torben G. Andersen and Francis X. Diebold. We would also like to thank the editor, Franz C. Palm, two anonymous referees, George Tauchen, Pete Kyle, Merlise Clyde, and seminar participants at the Duke Financial Econometrics Lunch Group and the Board of Governors in Washington, DC, for many useful comments which have significantly improved the paper.

References

Andersen, T.G., 1991. Estimation of systematic risk in the presence of non-trading: comments and extensions, Working paper, J.L. Kellogg Graduate School of Management, Northwestern University.

Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon. Journal of Empirical Finance 6, 457–477.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Market microstructure effects and the estimation of integrated volatility, Work in progress, Northwestern University, Duke University, and University of Pennsylvania.

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

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2003a forthcoming. Parametric and nonparametric volatility measurements. In: Hansen, L.P., Aït Sahalia, Y. (Eds.), Handbook of Financial Econometrics. North-Holland, Amsterdam.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003b. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

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.

Bai, X., Russell, J.R., Tiao, G.C., 2000. Beyond merton's utopia: effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data, Working paper, Graduate School of Business, University of Chicago.

Ball, R., Kothari, S.P., 1989. Nonstationary expected returns: implications for tests of market efficiency and serial correlation in returns. Journal of Financial Economics 25, 51–74.

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

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

Barndorff-Nielson, O.E., Shephard, N., 2002b. Estimating quadratic variation using realized variance. Journal of Applied Econometrics 17, 457–477.

Barucci, E., Reno, R., 2002. On measuring volatility and the GARCH forecasting performance. Journal of International Financial Markets, Institutions & Money 12, 183–200.

Berk, J.B., 2000. Sorting out sorts. Journal of Finance 55, 407–427.

Bollerslev, T., Melvin, M., 1994. Bid-ask spreads in the foreign exchange market: an empirical analysis. Journal of International Economics 36, 355–372.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96, 116–131.

Braun, P.A., Nelson, D.B., Sunier, A.M., 1995. Good news, bad news, volatility, and betas. Journal of Finance 50, 1575–1603.

Brennan, M.J., Wang, A.W., Xia, Y., 2002. Intertemporal asset pricing and the Fama–French three-factor model, Working paper, Department of Finance, UCLA.
Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton Univ. Press, Princeton, NJ.
Cho, Y.H., Engle, R.F., 1999. Time-varying betas and asymmetric effect of news: empirical analysis of blue chip stocks. NBER Working Paper 7330.
Christoffersen, P.F., Diebold, F.X., 1997. Optimal prediction under asymmetric loss. Econometric Theory 13, 808–817.
Cochrane, J.H., 2001. Asset Pricing. Princeton Univ. Press, Princeton, NJ.
Comte, F., Renault, E., 1998. Long-memory in continuous-time stochastic volatility models. Mathematical Finance 8, 291–323.
Corsi, F., Zumbach, G., Müller, U.A., Dacorogna, M.M., 2001. Consistent high-precision volatility from high-frequency data. Economic Notes 30, 183–204.
Dacorogna, M.M., Gencay, R., Müller, U.A., Olsen, R.B., Pictet, O.V., 2001. An Introduction to High-Frequency Finance. Academic Press, San Diego, CA.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1–33.
Davis, J.L., Fama, E.F., French, K.F., 2000. Characteristics, covariances, and average returns: 1929 to 1997. Journal of Finance 55, 389–406.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.
Epps, T.W., 1979. Co-movements in stock prices in the very short run. Journal of the American Statistical Association 74, 291–298.
Fama, E.F., French, K.F., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.
Fama, E.F., French, K.F., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
Fama, E.F., French, K.F., 1995. Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131–155.
Fama, E.F., French, K.F., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.
Fama, E.F., French, K.F., 1998. Value versus growth: the international evidence. Journal of Finance 53, 1975–1999.
Fama, E.F., MacBeth, J.D., 1973. Risk return and equilibrium: empirical tests. Journal of Political Economy 71, 607–636.
Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54, 1325–1360.
Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.
Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using realized volatility. Journal of Financial Economics 67, 473–509.
Foster, D., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample estimators. Econometrica 64, 139–174.
Ghysels, E., 1998. On stable factor structure in the pricing of risk: do time-varying betas help or hurt? Journal of Finance 53, 549–573.
Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross-section of returns. Journal of Political Economy (in press).
Hansen, L.P., Richard, S., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica 55, 587–613.
Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.
Hasbrouck, J., 1999. Security bid/ask dynamics with discreteness and clustering. Journal of Financial Markets 2, 1–28.
Hausman, J., Lo, A.W., MacKinlay, C., 1992. An ordered probit analysis of transaction prices. Journal of Financial Economics 31, 319–379.

Hodrick, R.J., Zhang, X., 2001. Evaluating the specification errors of asset pricing models. Journal of Financial Economics 62, 327–376.

Kothari, S.P., Shanken, J., Sloan, R.G., 1995. Another look at the cross-section of expected returns. Journal of Finance 50, 185–224.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Liew, J., Vassalou, M.G., 2000. Can book-to-market, size and momentum be risk factors that predict economic growth? Journal of Financial Economics 57, 221–245.

Lundin, M., Dacorogna, M.M., Müller, U.A., 1999. Correlation of high-frequency financial time series. In: Lequeux, P. (Ed.), Financial Markets Tick-by-Tick. Wiley, London, pp. 91–126.

Meddahi, N., 2002. A theoretical comparison between integrated and realized volatilities. Journal of Applied Econometrics 17, 475–508.

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

Nelson, D.B., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.

Orcutt, G.H., Winokur, H.S., 1969. First order autocorrelation: inference, estimation, and prediction. Econometrica 37, 1–14.

Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.

Taylor, S.J., Xu, X., 1997. The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance 4, 317–340.

Wang, X., 2000. Size effect, book-to-market effect, and survival. Journal of Multinational Financial Management 10, 257–273.

Zhang, B.Y.B., 2002. Time-varying volatilities, CAPM betas, and factor loadings: a high-frequency data perspective. PhD thesis, in preparation, Department of Economics, Duke University.

Zhou, B., 1986. High-frequency data and volatility in foreign exchange rates. Journal of Business and Economic Statistics 14, 45–52.

Zumbach, G., Corsi, F., Trapletti, A., 2002. Efficient estimation of volatility using high-frequency data. Manuscript, Olsen and Associates, Research Institute for Applied Economics, Zürich, Switzerland.