Do Implied Volatilities Predict Stock Returns?

Manuel Ammann, Michael Verhofen and Stephan Süss*

University of St. Gallen

Abstract

Using a complete sample of US equity options, we find a positive, highly significant relation between stock returns and lagged implied volatilities. The results are robust after controlling for a number of factors such as firm size, market value, analyst recommendations and different levels of implied volatility. Lagged historical volatility is - in contrast to the corresponding implied volatility - not relevant for stock returns. We find considerable time variation in the relation between lagged implied volatility and stock returns.

Keywords: Implied Volatility, Expected Returns

JEL classification: G10

*Manuel Ammann (manuel.ammann@unisg.ch) is professor of finance at the University of St. Gallen, Switzerland (Rosenbergstrasse 52, CH-9000 St. Gallen, Phone: +41 71 224 7000), Michael Verhofen (verhofen@gmail.com) is portfolio manager at Allianz Global Investors (Mainzer Landstrasse 11-13, D-60329 Frankfurt, Phone: +49 69 263 14394) and lecturer in finance at the University of St. Gallen, Switzerland, and Stephan Süss (stephan.suess@unisg.ch) is research assistant at the University of St. Gallen, Switzerland (Rosenbergstrasse 52, CH-9000 St. Gallen, Phone: +41 71 224 7000). We thank Sebastien Betermier, Peter Feldhütter, Thomas Gilbert, Sara Holland, Peter Tind Larsen, Miguel Palacios, Hari Phatak, David Skovmand and Ryan Stever for helpful comments.

1 Introduction

The option market reveals important information about investors' expectations of the underlying's return distribution. While considerable research has examined the informational content of index options, little is known about individual equity options. Using a complete sample of US equity options, we analyze the relation between implied volatility and future realized returns.

In the last three decades, several articles have documented a small degree of predictability in stock returns based on prior information, specifically at long horizons. In the long run, dividend yields on an aggregate stock portfolios predict returns with some success, as shown by Campbell & Shiller (1988), Fama & French (1988, 1989), as well as Goyal & Welch (2003). Additional variables found to have predictive power include the short-term interest rate (Fama & Schwert (1977)), spreads between long- and short-term interest rates (Campbell (1987)), stock market volatility (French, Schwert & Stambaugh (1987)), book-to-market ratios (Kothari & Shanken (1997), Pontiff & Schall (1998)), dividend-payout and price-earnings ratios (Lamont (1998)), as well as measures related to analysts' forecasts (Lee, Myers & Swaminathan (1999)). Baker & Wurgler (2000) detect a negative relationship between IPO activity and future excess returns. Lettau & Ludvigson (2001) find evidence for predictability using a consumption wealth ratio.

Recently, the relation between historical volatility and stock returns has been addressed by a number of authors (e.g., Goyal & Santa-Clara (2003), Bali, Cakici, Yan & Zhang (2005), and Ang, Hodrick, Xing & Zhang (2006)). Goyal & Santa-Clara (2003) analyze the predictability of stock market returns with several risk measures. They find a significant positive relation between the cross-sectional average stock variance and the return on the market, whereas the variance of the market has no forecasting power for the market return. However, Bali et al. (2005) disagree with these findings.

They argue that the results are primarily driven by small stocks traded on the NASDAQ, and are therefore partially due to a liquidity premium. Moreover, the results do not hold for an extended sample period. Ang et al. (2006) examine the pricing of aggregate volatility risk in the cross-section of stock returns. They find that stocks with high idiosyncratic volatility in the Fama and French three-factor model have very low average returns.

Option-implied volatility is different from most of the variables used for predicting stock returns in at least two respects. First, it is a real forward-looking variable measuring market participants' expectations. Second, it is a traded price and therefore less likely to be affected by biases.

To our best knowledge, no study exists that systematically analyzes the informational content of implied volatility in the cross-section. Existing studies have only focussed on index option data or a very small sample of single equity options. This study contributes to the existing literature by investigating the relation between implied volatility and stock returns on a very large data basis.

To analyze the relation between implied volatility and stock returns, we apply a predictive regression approach in univariate and multivariate settings. Our results are based on the OptionMetrics database, which contains a survivial bias-free, complete data set of implied volatilities for the US stock market. To control for a number of factors and to investigate the stability of the findings, we merge our sample with the CRSP, Compustat, and IBES FirstCall data. Model misspecification is addressed by using different regression settings. We address parameter uncertainty by a bootstrapping and an additional rolling-windows approach.

We find a highly significant, positive relation between returns and lagged implied volatilities. This dependence is stronger for firms with small market capitalizations and is independent of different valuation levels, measured by the book-to-market ratio. Our findings are persistent after controlling for

market risk (using the CAPM) and the exposure to the risk factors in the Carhart four-factor model. The informational content of first-order differences of implied volatility seems to be limited. With respect to analyst recommendations, we find weaker relations between returns and lagged implied volatilities for companies with higher analyst coverages. The findings seem to be stable for different times to maturity of implied volatility. Historical volatilities do not seem to have the same informational content as implied volatilities. We find considerable time variation in the relation between lagged implied volatility and stock returns. The out-of-sample predictive power is weak compared to the iid model.

The paper is organized as follows: Section 2 outlines our research design, as well as the applied data set. Section 3 presents the empirical results. Section 4 concludes.

2 Research Design

2.1 Data

To obtain the data set for our empirical analysis, we merge five different databases. From OptionMetrics, we retrieve option price data and historical volatilities. Equity return data are obtained from the CRSP database. The book values and market capitalization figures are from Compustat. Analyst forecasts are collected from IBES. The respective risk premia are from the Fama and French database. Data are merged by the respective CUSIP as identifier. Our sample with monthly frequency covers the period from January 1996 to December 2005.

The OptionMetrics database is described in detail in Optionmetrics (2005). Our study is based on implied volatility for standardized call options with a maturity of 91 calendar days and a strike price equal to the forward price. They are computed as outlined in Optionmetrics (2005). In addittion to

implied volatilities, historical volatilities are retrieved from OptionMetrics. For comparability reasons, historical volatility is also computed over a time period of 91 calendar days.

To account for systematic risk, we use the risk factors of the Fama & French (1993) and Carhart (1997) models. The data for the market portfolio (MRP), the "high-minus-low"(HML), the "small-minus-big"(SMB), the momentum factor (UMD), and the risk-free interest rate (RF) are from the Fama and French data library.

2.2 Predictive Regressions and Panel Data

Predictive regressions (e.g., Fama & French (1989), Stambaugh (1999)) regress future returns on predictive variables or, equivalently, returns, r_t, on lagged predictive variables, $x_{t-1,i}$,

$$r_t = \alpha + \beta_1 x_{t-1,1} + \beta_2 x_{t-1,2} + ... + \beta_k x_{t-1,k} + ... + \beta_K x_{t-1,K} + \varepsilon_t, \quad (1)$$

where t denotes the time index, k the index for the K predictive variables, α the constant, β_k the respective factor loading, and ε_t the error term. However, equation (1) is only applicable for the single-asset case. In the case of K assets, a panel data approach can be used.

The error representation for the linear fixed-effect panel data model is (Frees (2004))

$$r_{it} = \alpha_i + \beta_1 x_{it,1} + \beta_2 x_{it,2} + ... + \beta_K x_{it,K} + \varepsilon_{\iota t}, \quad (2)$$

where $\mathbb{E}(\varepsilon_{\iota t}) = 0$. The parameters β_j are common to each subject and called global (or population parameters). The parameters α_i vary by subject and are known as individual, or subject-specific, parameters.

To analyze the relation between implied volatility and returns, we regress

returns on lagged implied volatilities, $\mathcal{V}$,

$$r_{it} = \alpha_i + \beta_1 \mathcal{V}_{it-1,1} + \varepsilon_{\iota t} \quad . \tag{3}$$

The estimated factor loading β_1 therefore summarizes the full sample relation between implied volatility and future stock returns.

2.3 Excess Returns

Besides raw returns, we use the CAPM and the Carhart (1997) four-factor model to account for systematic risk effects. To estimate the exposure towards the Fama & French (1993) risk factors and the Carhart (1997) momentum factor, we run the following regression for each asset i to control for market, size, value, and momentum risk

$$\begin{aligned} r_{it} - r_{ft} &= \alpha_{i,FF} + \beta_{i,MRP} \cdot MRP_t + \beta_{i,HML} \cdot HML_t + \\ &\quad \beta_{i,SMB} \cdot SMB_t + \beta_{i,UMD} \cdot UMD_t + \varepsilon_{it} \end{aligned} \tag{4}$$

and the following regression to control for market risk

$$r_{it} - r_{ft} = \alpha_{i,CAPM} + \beta_{i,MRP} \cdot MRP_t + \varepsilon_{it} \quad . \tag{5}$$

2.4 Robustness

To analyze the robustness of our findings, we perform a number of different analyses. First, we run the respective regressions for various subsamples. Second, we use a rolling-window approach to account for time-varying factor loadings. Third, we implement a bootstrapping approach to investigate possible problem with the estimated used. Finally, we analyze the out-of-sample performance.

To analyze the out-of-sample validity of the models we regress the realizations of each return r_{it} on the corresponding time-$t-1$ return forecast $\hat{r}_{it-1}$,

i.e.,

$$r_{it} = \alpha + \beta \cdot \hat{r}_{it,t-1} + \varepsilon_{\iota t}. \tag{6}$$

Under accurate forecasts, we expect $\alpha = 0$ and $\beta = 1$.

3 Empirical Results

3.1 Regressions of Returns on Lagged Implied Volatility

Table 1 shows the basic results of this paper. In the first column, we show the estimated factor loadings from a regression of returns on lagged implied volatility. An estimated factor loading of 2.021 indicates that a 1% higher implied volatility leads, on average, to a return increase of 2.021% in the subsequent month. This finding is highly significant with a t-value of 9.457. The goodness-of-fit of this model, measured by R^2, is 0.8%.

The second column illustrates the estimated factor loading from a simple iid model. Under the assumption of no predictability in returns, the best forecast is a constant. The root-mean-squared-error (RMSE) of the iid model is 16.8408. This value is only marginally higher than the RMSE value of 16.8379 for the model with implied volatility. Since these two values are very similar, the findings suggest that the degree of predictability is low even though the estimated factor loadings are highly significant.

To test for nonlinearity, we include the squared implied volatility in the regression equation. The results in the third column show that the estimated coefficient is insignificant with a value of 0.076. This suggests the appropriateness of the linear model.

To account for time-varying means and dispersion of implied volatility, we compute standardized z-scores. The regression of returns on standardized, lagged implied volatility validates previous findings. With an estimated coefficient of 0.939, we find a highly significant, positive relation between implied volatility and future returns with a factor loading of 0.032.

To account for time-varying means and dispersion in stocks returns, we also compute standardized z-scores for every month for the return data. A regression of standardized returns on standardized, lagged implied volatility reveals, as before, a positive and significant relation between risk and return. To analyze the robustness of these findings, we perform a number of different analyses. First, we investigate whether the relation between returns and lagged implied volatility is also valid for different levels of implied volatility. For example, stocks with high implied volatility might behave differently than stocks with comparably lower implied volatility.

Table 2 shows the estimated factor loadings for different subsamples. We re-estimate the forecasting model for stocks with an implied volatility between 0% and 20% (subsample 1), 20% and 40% (subsample 2), 40% and 60% (subsample 3), 60% to 80% (subsample 4), and 80% to 100% (subsample 5). We find a positive, highly significant relation between returns and lagged implied volatilities for subsamples 1, 3, 4, and 5. The estimated factor loadings are of comparable magnitude for subsamples 1 and 3 (6.559 and 7.755) and for subsamples 4 and 5 (13.525 and 12.417). However, the findings for subsample 2 are different. The estimated factor loading with a value of -0.825 is slightly negative, but insigniicant.

3.2 Size and Value Effects

Table 3 outlines the estimated factor loadings for separate regressions for different quintiles of market capitalizations and book-to-market ratios.

With respect to market capitalization, we find that the strength of the relation between anticipated risk and the subsequent return decreases with higher market values. For stocks with the highest market capitalization (Q5), we estimate a factor loading of 3.190 while the factor loading for small stocks, e.g., in quantile 2 (Q2), is 9.704. All findings are highly significant. The factor loading for growth stocks (Q1) is, with a value of 6.344, very

similar to the corresponding factor loading of value stocks (Q5), which has a value of 7.097.

3.3 Excess Returns

Table 4 shows the results of the regression of excess returns on lagged implied volatilities for the full sample and for various subsamples. In the upper part of the table, excess returns have been computed against the CAPM model and in the lower part against the Carhart four-factor model. Subsamples are formed on different levels of implied volatility.

The estimated factor loadings are positive and highly significant for all samples. While the factor loading of implied volatility is 2.021 for raw returns (see Table 1), it is higher when controlling for systematic risk factors. Against the CAPM, the coefficient is 7.772 for the full sample, against the Carhart four-factor model, the corresponding coefficient has a value of 6.408. Both factor loadings are highly significant. We conclude that implied volatility carries some information beyond that implied by the CAPM and the Carhart four-factor model.

For subsamples formed on different levels of implied volatility, we find that, in general, factor loadings increase with higher levels of implied volatility. The subsample regressions validate the findings for the full sample.

3.4 First-Order Differences

Table 5 illustrates the estimated factor loadings of a regression of returns on lagged, first-order differences of implied volatility. The first column shows the results for the full sample, the remaining columns the respective results of the regressions for various subsamples formed on the magnitude of first-order differences of implied volatility.

With a value of 2.648 for the full sample, we find a highly significant, positive relation between the returns and the lagged change in implied volatility.

Therefore, an investor can expect a higher monthly return for a stock if implied volatility has increased in the previous month.
For subsamples formed on different directions and magnitudes of the change in implied volatility, the results differ. First, we find hardly any significance between the change in implied volatility and future returns. Second, the estimated factor loadings differ substantially for different subsamples.

3.5 Analyst Forecasts

Table 6 shows the estimated factor loadings for subsamples formed on the mean analyst recommendation. Quantile 1 contains the most favorable recommendations, Quantile 5 the least favorable recommendations. The general observation, i.e., the positive relation between returns and lagged implied volatility, holds for all subsamples based on different levels of analyst recommendations. The findings are highly significant in all cases. The magnitude of the relation between implied volatility and returns differs slightly for different levels of analyst recommendations. For stocks with very positive (Q1) or very negative recommendations (Q5), the estimated factor loadings of 6.855 and 8.866, respectively, are higher than for stocks with an average recommendation (Q3) where the value is 5.003.
Table 6 also shows the estimated regression coefficients for subsamples formed on the number of analysts covering a specific stock. For all subsamples, the relation between returns and lagged implied volatility is positive on a high significance level. However, we find a monotonic decreasing relation between the estimated coefficients and the number of recommendations. The higher the number of analysts following a particular stock, the lower the informational content of implied volatility.

3.6 Implied Volatility vs. Historical Volatility

In Table 7, we outline the results of various regressions of returns on different lagged variables. We analyze the dependence between returns and implied, as well as historical volatilities with time horizons of 30, 60 and 91 days. We find a very clear pattern. For all three different maturities of implied volatilities, the estimated coefficients are highly significant and, with values between 1.734 and 2.021, very similar. In contrast, we do not find a similar pattern for historical volatility. The estimated coefficients are significant at a 0.2% level for a time horizon of 30 days, and on a 5% level for a time horizon of 91 days, but not for a time horizon of 60 days.

3.7 Univariate Regressions

Table 1 shows the histogram for the univariate regression of returns on lagged implied volatility. The results should be interpreted carefully. Due to the small sample size (monthly data for a maximum of 9 years), not all coefficients are significant. Two main findings can be seen in Figure 2. First, there is considerable cross-sectional dispersion in the factor loadings. Second, the relation between implied volatility and return is, on average, positive.

3.8 Parameter Uncertainty and Bootstrapping

Table 8 illustrates the estimated factor loadings of separate regressions for each full year in the sample period, i.e. from 1996 to 2005. The estimated factor loading of lagged implied volatility on return varies between 7.333 in 2003 and 33.482 in 2001. Therefore, there is always a positive, highly significant relation between perceived risk and the subsequent return.
The goodness-of-fit varies substantially over time. In 2000 and 2001, the model can explain more than 2% of total variance (2.20% and 2.33%). In other years, e.g. 1996, 1999, and 2005, the R^2 was very low, taking values

between 0.21% and 0.28%.

Figure 2 shows the estimated factor loading for a rolling window of 60 months (5 years). Roughly speaking, the graph indicates an increase in the factor loading from 4 to 11 from 2000 until 2002. In 2003, the factor loading dropped to 3 and fluctuated between 2 and 6 until the end of the sample period. Therefore, we find considerable time variation in the magnitude of the relation between implied volatility and return. However, the estimated factor loading is positive at any point in time.

Figure 3 shows the histogram of bootstrapped factor loadings for the full sample regression and Figure 4 the associated t-values. As shown in Table 1, the full-sample estimated coefficient is 2.021. Its corresponding t-value is 9.457. Both figures indicate that the findings are not spurious.

3.9 Out-of-Sample Performance

Table 9 gives the results from a predictive regression for the fixed effects panel data model and the iid model. The parameters for both models are estimated over a rolling horizon of 60 months. Based on the estimated parameters, a return is predicted for the next month. The realized returns are regressed on their corresponding predictions.

If forecasts are perfect, we expect a constant of 0 and slope coefficients of 1. However, the empirical findings are quite different. For the one factor model with implied volatility as predictive variable, the estimated constant is -1.169 and the slope coefficient has a value of -0.262. For a naive, iid model, the estimated constant is also -1.169 and the slope coefficient is -0.293.

In its last row, Table 9 shows that the $RMSE$ of the one-factor model is, with a value of 17.373, higher than for the iid model with a value of 16.080.

4 Conclusion

To analyze the relation between implied volatility and stocks returns, we use a predictive-regression approach in an univariate and multivariate setting. We use the OptionMetrics database, which contains a survivial bias-free, complete database for implied volatilities for the US stock market. A merge of the database with CRSP, Compustat, and IBES FirstCall data allows to control for a number of factors and to investigate the stability of the findings. Model misspecification is evaluated by using different regression settings. Parameter uncertainty is addressed by a bootstrapping approach and a rolling windows approach. Furthermore, we consider the out-of-sample validity.

As our main finding, we observe a highly significant, positive relation between returns and lagged implied volatilities. This relation is weaker for larger market capitalizations and independent of different valuation levels (using the book-to-market ratio). These findings are persistent after controlling for market risk (using the CAPM) and the risk factors of the Carhart four-factor model. The informational content of first-order differences of implied volatilities seems to be limited. With respect to analyst recommendations, we find weaker relations between returns and lagged implied volatilities for companies with high analyst coverages. A comparison of implied volatilities for different time horizons shows that the patterns seem to be stable for different time to maturities. Historical volatilities do not carry the same informational content as implied volatilities. We find considerable time variation in the relation between lagged implied volatility and stock returns. The out-of-sample predictive power is weak compared to the iid model.

References

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2006), 'The cross-section of volatility and expected returns', *The Journal of Finance* **61**, 259.

Baker, M. & Wurgler, J. (2000), 'The equity share in new issues and aggregate stock returns', *Journal of Finance* **55**, 2219–2257.

Bali, T. G., Cakici, N., Yan, X. & Zhang, Z. (2005), 'Does idiosyncratic risk really matter?', *Journal of Finance* **60**, 905–929.

Campbell, J. (1987), 'Stock returns and the term structure', *Journal of Financial Economics* **18**, 373–399.

Campbell, J. & Shiller, R. (1988), 'The dividend-price ratio and expectations of future dividends and discount factors', *Review of Financial Studies* **1**, 195–227.

Carhart, M. (1997), 'On persistence in mutual fund performance', *Journal of Finance* **52**, 57–82.

Fama, E. & French, K. (1989), 'Business conditions and expected returns on stocks and bonds', *Journal of Financial Economics* **19**, 3–29.

Fama, E. & French, K. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**, 3–57.

Fama, E. & Schwert, G. (1977), 'Asset returns and inflation', *Journal of Financial Economics* **5**, 115–146.

Frees, E. W. (2004), *Longitudinal and Panel Data*, Cambridge University Press, Cambridge.

French, K., Schwert, G. & Stambaugh, R. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 293–305.

Goyal, A. & Santa-Clara, P. (2003), 'Idiosyncratic risk matters!', *Journal of Finance* **58**, 975–1007.

Goyal, A. & Welch, I. (2003), 'Predicting the equity premium with dividend ratios', *Management Science* **49**, 639–654.

Kothari, S. & Shanken, J. (1997), 'Book-to-market, dividend yield, and expected market returns: A time series analysis', *Journal of Financial Economics* **44**, 169–203.

Lamont, O. (1998), 'Earnings and expected returns', *Journal of Finance* **53**, 1563–1587.

Lee, C., Myers, J. & Swaminathan, B. (1999), 'What is the intrinsic value of the Dow', *Journal of Finance* **54**, 1639–1742.

Lettau, M. & Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *Journal of Finance* **56**, 815–849.

Optionmetrics (2005), 'Ivy DB: File and data reference manual, version 2.5'.

Pontiff, J. & Schall, L. (1998), 'Book-to-market as a predictor of market returns', *Journal of Financial Economics* **49**, 141–160.

Stambaugh, R. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**, 315–421.

Table 1: Regressions of Returns on Lagged Implied Volatility

This table illustrates the estimated coefficients, t-values (in parentheses), sample sizes, root-mean-squared-errors, as well as the respective R^2 for fixed-effects panel data regression of returns on lagged implied volatilities for the full sample. To test for non-linearities, we use standardized and squared implied volatilities and returns. $z(.)$ denotes a standardized variable with zero mean and unit variance. The results are based on the merged CRSP, Compustat, IBES, and OptionMetrics databases with monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$z(r_{t,t+1})$
Subsample	Full	Full	Full	Full	Full
$\mathcal{V}_{91}$	2.021†††		1.902†††		
	(9.457)		(3.953)		
$(\mathcal{V}_{91})^2$			0.076		
			(0.277)		
$z(\mathcal{V}_{91})$				0.939†††	0.032†††
				(14.955)	(9.673)
$(z(\mathcal{V}_{91}))^2$					
constant	-1.474†††	-0.463†††	-1.438†††	-0.465†††	-0.000
	(-13.170)	(-13.970)	(-8.466)	(-14.038)	(-0.004)
N	258281	258281	258281	258281	258281
$RMSE$	16.8379	16.8408	16.8379	16.8334	0.8756
R^2	0.0080	0.0000	0.0080	0.0055	0.0061

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 2: Different Levels of Implied Volatility

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged implied volatilities for various subsamples based on different levels of implied volatilities. The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	$0.0 \leq \mathcal{V}_{91} < 0.2$	$0.2 \leq \mathcal{V}_{91} < 0.4$	$0.4 \leq \mathcal{V}_{91} < 0.6$	$0.6 \leq \mathcal{V}_{91} < 0.8$	$0.8 \leq \mathcal{V}_{91} < 1.0$
$\mathcal{V}_{91}$	6.559††	-0.825	7.755†††	13.525†††	12.417†††
	(3.155)	(-1.279)	(6.928)	(6.834)	(3.345)
constant	-0.268	0.828†††	-3.992†††	-10.529†††	-13.348†††
	(-0.779)	(4.167)	(-7.231)	(-7.699)	(-4.054)
N	14765	96726	72494	41537	20131
$RMSE$	4.9745	9.1865	14.8816	20.7627	26.7389
R^2	0.0002	0.0003	0.0005	0.0001	0.0002

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 3: Size and Value Effects

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged implied volatilities various subsamples based on different levels of market value (MV) and the book-to-market (BTM) ratio. Q denotes the quantile. Stocks with a high book-to-market ratio, e.g., in Q5, can be interpreted as value stocks and stocks with a low book-to-market ratio, e.g., in Q1, as growth stocks. The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	MV in Q1	MV in Q2	MV in Q3	MV in Q4	MV in Q5
$\mathcal{V}_{91}$	40.250†††	9.704†††	7.714†††	6.904†††	3.190†††
	(5.066)	(6.354)	(11.392)	(15.898)	(10.232)
constant	-24.918†††	-7.564†††	-6.023†††	-4.542†††	-1.349†††
	(-5.162)	(-6.765)	(-13.003)	(-18.302)	(-10.100)
N	501	7140	33789	77327	126994
$RMSE$	19.2760	19.6963	20.0507	18.1044	14.1366
R^2	0.0079	0.0032	0.0036	0.0070	0.0102
Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	BTM in Q1	BTM in Q2	BTM in Q3	BTM in Q4	BTM in Q5
$\mathcal{V}_{91}$	6.344†††	7.277†††	7.437†††	10.903†††	7.097†††
	(12.661)	(15.379)	(14.320)	(17.168)	(8.854)
constant	-5.042†††	-3.970†††	-3.300†††	-4.421†††	-3.199†††
	(-16.710)	(-16.540)	(-13.738)	(-15.140)	(-7.652)
N	69862	68454	50501	35343	18005
$RMSE$	19.9776	16.1168	14.1463	14.0702	15.1247
R^2	0.0099	0.0055	0.0066	0.0018	0.0032

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 4: Excess Returns

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of excess returns on lagged implied volatilities for the full sample and various subsamples. Excess returns have been computed using univariate OLS regression with a CAPM model and with a Carhart four-factor model. The data for the risk premia are from Fama and French. The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}^{(CAPM)}$	$r_{t,t+1}^{(CAPM)}$	$r_{t,t+1}^{(CAPM)}$	$r_{t,t+1}^{(CAPM)}$	$r_{t,t+1}^{(CAPM)}$	$r_{t,t+1}^{(CAPM)}$
Subsample	Full	$0.0 \leq \mathcal{V}_{91} < 0.2$	$0.2 \leq \mathcal{V}_{91} < 0.4$	$0.4 \leq \mathcal{V}_{91} < 0.6$	$0.6 \leq \mathcal{V}_{91} < 0.8$	$0.8 \leq \mathcal{V}_{91} < 1.0$
$\mathcal{V}_{91}$	7.772†††	5.181††	2.002†††	8.593†††	15.645†††	16.548†††
	(40.871)	(2.679)	(3.457)	(8.574)	(8.893)	(4.924)
constant	-3.495†††	-0.300	-0.150	-4.065†††	-10.656†††	-13.753†††
	(-35.097)	(-0.940)	(-0.838)	(-8.224)	(-8.765)	(-4.613)
N	258281	14765	96726	72494	41537	20131
$RMSE$	14.9835	4.6272	8.2457	13.3243	18.4576	24.2098
R^2	0.0001	0.0000	0.0001	0.0002	0.0001	0.0001
Dependent	$r_{t,t+1}^{(Carhart)}$	$r_{t,t+1}^{(Carhart)}$	$r_{t,t+1}^{(Carhart)}$	$r_{t,t+1}^{(Carhart)}$	$r_{t,t+1}^{(Carhart)}$	$r_{t,t+1}^{(Carhart)}$
Subsample	Full	$0.0 \leq \mathcal{V}_{91} < 0.2$	$0.2 \leq \mathcal{V}_{91} < 0.4$	$0.4 \leq \mathcal{V}_{91} < 0.6$	$0.6 \leq \mathcal{V}_{91} < 0.8$	$0.8 \leq \mathcal{V}_{91} < 1.0$
$\mathcal{V}_{91}$	6.408†††	6.942†††	4.453†††	6.961†††	11.758†††	13.681†††
	(36.265)	(3.649)	(8.070)	(7.324)	(7.112)	(4.427)
constant	-2.692†††	-1.061†††	-1.101†††	-3.060†††	-7.474†††	-10.588†††
	(-29.094)	(-3.374)	(-6.474)	(-6.528)	(-6.541)	(-3.862)
N	258192	14744	96691	72470	41530	20129
$RMSE$	13.9217	4.5507	7.8582	12.6348	17.3460	22.2638
R^2	0.0010	0.0004	0.0001	0.0000	0.0001	0.0002

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 5: First-Order Differences of Implied Volatilities

The table shows the estimated coefficients, the t-values (in parenthesis), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged, first-order difference of implied volatilities for the full sample and various subsamples based on the magnitude of the change in implied volatilities. The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	Full	$\Delta\mathcal{V}_{91} <= -0.100$	$\Delta\mathcal{V}_{91} <= -0.050$	$\Delta\mathcal{V}_{91} <= -0.025$	$\Delta\mathcal{V}_{91} <= 0.000$
			$\Delta\mathcal{V}_{91} > -0.100$	$\Delta\mathcal{V}_{91} > -0.050$	$\Delta\mathcal{V}_{91} > -0.025$
$\Delta\mathcal{V}_{91}$	2.648†††	3.524†	-3.055	-15.528	1.199
	(7.744)	(2.415)	(-0.373)	(-1.251)	(0.149)
constant	-0.451†††	-1.862†††	-1.045	-1.017†	0.092
	(-13.498)	(-5.802)	(-1.782)	(-2.235)	(0.854)
N	251964	17305	26692	31899	57099
$RMSE$	16.7753	22.4897	16.9880	14.5683	12.9712
R^2	0.0000	0.0011	0.0001	0.0000	0.0001

Dependent		$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample		$\Delta\mathcal{V}_{91} <= 0.025$	$\Delta\mathcal{V}_{91} <= 0.050$	$\Delta\mathcal{V}_{91} <= 0.0100$	$\Delta\mathcal{V}_{91} > 0.100$
		$\Delta\mathcal{V}_{91} > 0.000$	$\Delta\mathcal{V}_{91} > 0.025$	$\Delta\mathcal{V}_{91} > 0.050$	
$\Delta\mathcal{V}_{91}$		-7.085	6.324	-0.333	0.054
		(-0.817)	(0.423)	(-0.033)	(0.035)
constant		0.165	-0.352	-0.437	-1.554†††
		(1.439)	(-0.644)	(-0.600)	(-4.239)
N		51496	26278	22546	18649
$RMSE$		13.2949	15.6549	19.2214	26.7819
R^2		0.0001	0.0000	0.0003	0.0013

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 6: Impact of Analyst Recommendations

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged implied volatilities for various subsamples formed on the mean recommendation of analysts and on the total number of analyst recommendations. $MeanRec$ is a variable provided by IBES and takes values between 1 and 5 where 1 correspondents to -strong buy- and 5 to -strong sell-. Quantile 1 (Q1) to Quantile 5 (Q5) denote the quantile of the mean recommendation and the number of recommendations. Q1 contains the stocks with the highest average recommendation (upper part) and the lowest number of analysts (lower part). Q5 contains the stocks with the lowest average reommendation (upper part) and the highest number of analysts (lower part). The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	$MeanRec$ in Q1	$MeanRec$ in Q2	$MeanRec$ in Q3	$MeanRec$ in Q4	$MeanRec$ in Q5
$\mathcal{V}_{91}$	6.855†††	4.212†††	5.003†††	7.528†††	8.806†††
	(8.677)	(7.200)	(9.232)	(14.270)	(13.504)
constant	-5.091†††	-2.575†††	-2.329***	-3.431†††	-4.633†††
	(-11.184)	(-8.510)	(-8.933)	(-13.657)	(-13.507)
N	34205	53444	49894	53434	32483
$RMSE$	18.9688	16.9825	15.0278	15.0673	16.1858
R^2	0.0090	0.0064	0.0066	0.0038	0.0046
Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	$NumRec$ in Q1	$NumRec$ in Q2	$NumRec$ in Q3	$NumRec$ in Q4	$NumRec$ in Q5
$\mathcal{V}_{91}$	10.296†††	8.137†††	7.631†††	6.905†††	2.132†††
	(6.389)	(9.404)	(11.649)	(14.055)	(5.728)
constant	-6.891†††	-5.535†††	-4.964†††	-4.023†††	-0.780†††
	(-6.739)	(-10.355)	(-13.380)	(-15.616)	(-4.727)
N	7094	21912	38578	65476	90400
$RMSE$	19.0643	18.8151	17.6981	17.1366	14.3742
R^2	0.0059	0.0059	0.0043	0.0068	0.0060

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 7: Implied Volatilities vs. Historical Volatilities

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged implied volatilities and lagged historical volatilities with maturities of 30, 60 and 91 calendar days. The regressions indicate a robust, highly significant relation between returns and lagged implied volatilites, but not between returns and lagged historical volatilities. The results base on the merged CRSP, Compustat, IBES, and OptionMetrics databases. The data set consists of monthly data from January 1996 to December 2005. $\mathcal{V}_{30}$, $\mathcal{V}_{60}$, $\mathcal{V}_{91}$ are variables provided by OptionMetrics and contain the standardized implied volatility for at-the-money call equity options with a time to maturity of 30, 60 or 91 calendar days, respectively. $\mathcal{V}_{30}^{(hist)}$, $\mathcal{V}_{60}^{(hist)}$, $\mathcal{V}_{91}^{(hist)}$ are historical volatilities provided by OptionMetrics.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	Full	Full	Full	Full	Full	Full
$\mathcal{V}_{30}$	1.902†††					
	(9.900)					
$\mathcal{V}_{60}$		1.734†††				
		(8.622)				
$\mathcal{V}_{91}$			2.021†††			
			(9.457)			
$\mathcal{V}_{30}^{(hist)}$				0.506†††		
				(4.081)		
$\mathcal{V}_{60}^{(hist)}$					0.263	
					(1.867)	
$\mathcal{V}_{91}^{(hist)}$						0.337†
						(2.213)
constant	-1.428†††	-1.342†††	-1.474†††	-0.773†††	-0.683†††	-0.757†††
	(-13.782)	(-12.474)	(-13.170)	(-10.917)	(-8.530)	(-8.745)
N	261810	259739	258281	268406	264870	262813
$RMSE$	16.8218	16.8402	16.8379	17.5222	17.5830	17.6062
R^2	0.0075	0.0080	0.0080	0.0062	0.0078	0.0085

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 8: Time Varying Factor Loadings

The table shows the estimated coefficients, the t-values (in parentheses), the sample size and the goodness-of-fit from a fixed-effects panel data regression of returns on lagged implied volatilities for the full sample and various subsamples. The results base on the merged CRSP and OptionMetrics databases. The data set consists of monthly data from January 1996 to April 2006. $\mathcal{V}_{91}$ is a variable provided by OptionMetrics and contains the standardized implied volatility for at-the-money call equity options with a time to maturity of 91 calendar days.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	Year 1996	Year 1997	Year 1998	Year 1999	Year 2000
$\mathcal{V}_{91}$	12.209†††	15.413†††	30.218†††	15.785†††	21.546†††
	(12.239)	(14.839)	(36.533)	(15.612)	(18.271)
constant	-4.488†††	-6.549†††	-16.175†††	-8.631†††	-17.676†††
	(-10.300)	(-13.803)	(-37.098)	(-14.655)	(-20.835)
N	20965	25385	28761	29131	25988
$RMSE$	12.8450	13.8774	17.9397	17.0623	24.6935
R^2	0.0023	0.0088	0.0037	0.0021	0.0220
Dependent	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Subsample	Year 2001	Year 2002	Year 2003	Year 2004	Year 2005
$\mathcal{V}_{91}$	33.482†††	29.604†††	7.333†††	21.593†††	9.831†††
	(26.958)	(30.740)	(7.364)	(17.485)	(10.716)
constant	-23.513†††	-19.731†††	0.358	-8.305†††	-2.919†††
	(-30.398)	(-37.376)	(0.828)	(-17.362)	(-8.809)
N	24910	25429	24322	26653	26737
$RMSE$	20.6631	17.7816	11.9665	11.2352	10.7640
R^2	0.0233	0.0094	0.0052	0.0146	0.0028

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Table 9: Out-of-Sample Performance

The Table shows the the estimated coefficients of a regression of realized returns on forecasted returns for a linear, fixed effects model with implied volatility as a predictive variable and for the iid model. The out-of-sample forecast bases on a rolling window with a size of 60 months. $RMSE$ provides the root-mean-squared-error of the prediction error.

Dependent	$r_{t,t+1}$	$r_{t,t+1}$
$\hat{r}_{t,t+1}$	-0.262†††	
	(-36.048)	
$\hat{r}_{iid,t,t+1}$		-0.293†††
		(-18.788)
$constant$	-1.169†††	-1.169†††
	(-24.983)	(-21.416)
N	126857	126857
$RMSE$	17.375	16.080
R^2	0.0101	0.0028

† indicates significance on a 95.0%, †† on a 99.0% and ††† on a 99.9% level.

Figure 1: Histogram of Estimated Factor Loadings in an Univariate Regression

The figure shows the histogram of estimated factor loadings of an univariate regressions of returns on lagged implied volatility for each stock in the sample. On average over all stocks, there is a positive relation between lagged implied volatility and stock returns.

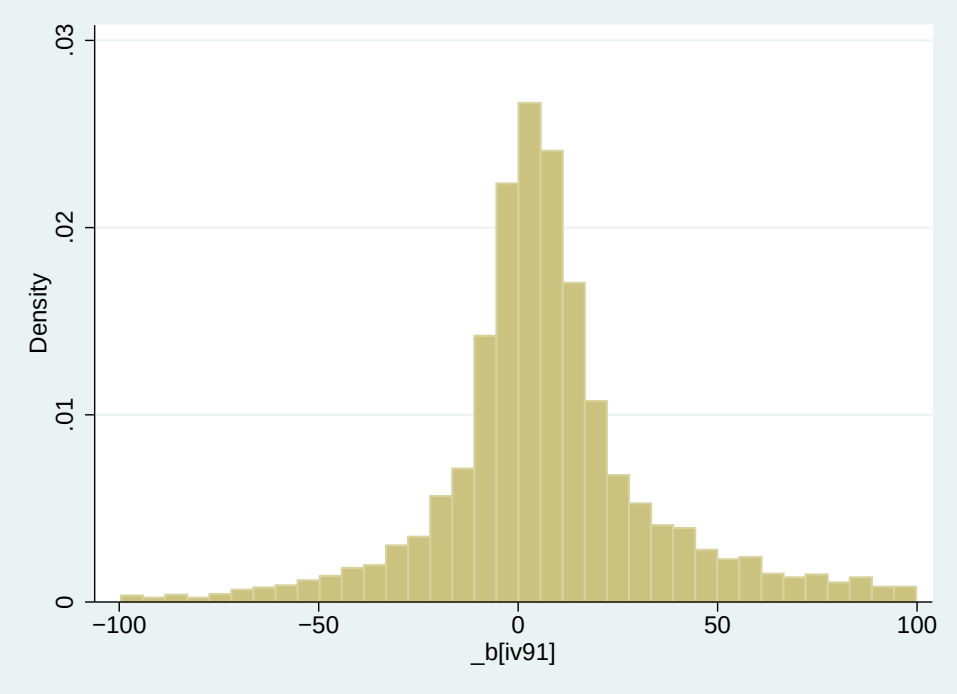


Figure 2: Factor Loadings in a Rolling Regression

The figure shows the estimated factor loading in a rolling, fixed effects panel data regression of returns on lagged implied volatility. The windows size is 60 months. The results indicate considerable time variation of the factor loading of implied volatility on returns.

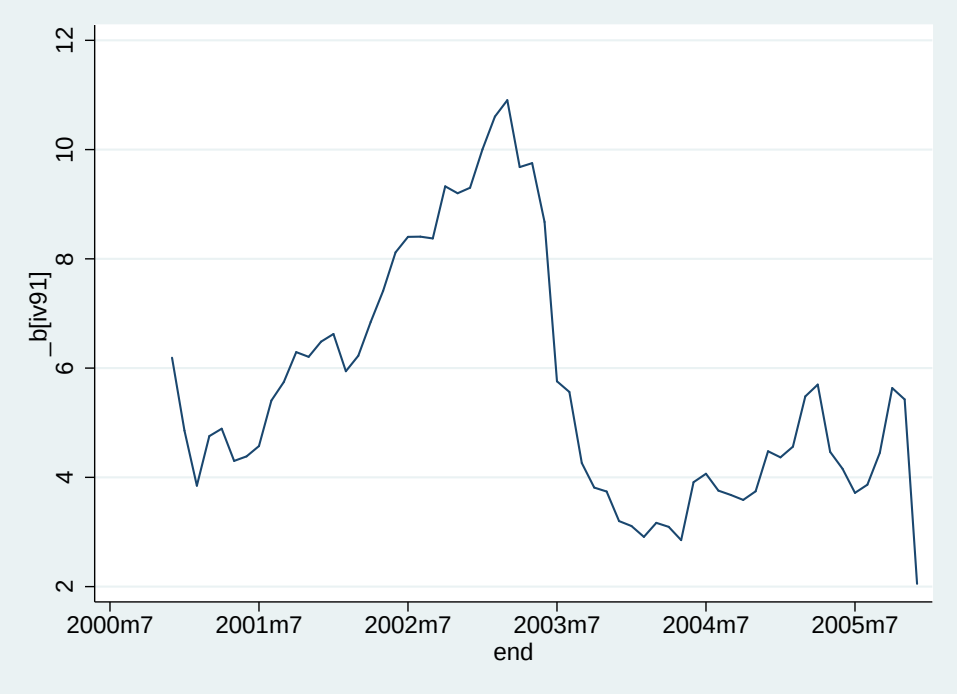


Figure 3: Histogram of Boostrapped Factor Loadings

The figure shows the bootstrapped factor loadings on implied volatility of a regression of returns on lagged implied volatilities in a fixed effects panel data regression. The results indicate that the factor loading of lagged implied volatility on stock returns is between 1.5 and 2.5.

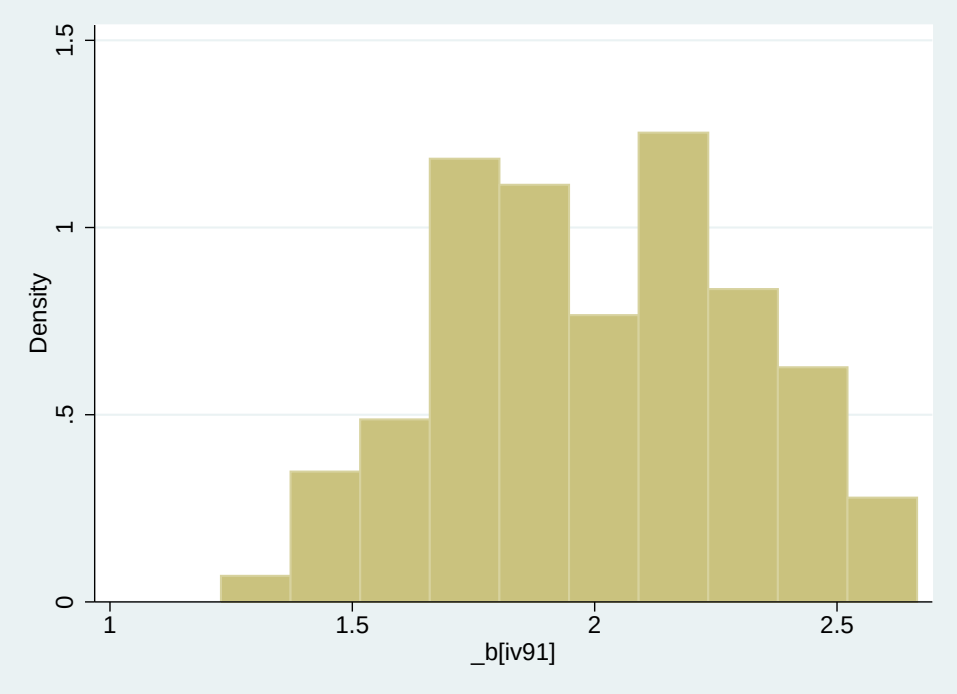


Figure 4: Histogram of Boostrapped T-Values

The figure shows the bootstrapped t-values for the factor loadings on implied volatility of a regression of returns on lagged implied volatilities in a fixed effects panel data regression. The results indicate that the estimated t-values are highly significant and robust.

