On the Relation between EGARCH Idiosyncratic Volatility and Expected Stock Returns

Hui Guo, Haimanot (Haim) Kassa, and Michael F. Ferguson*

First Draft: August 2010

This Draft: July 2012

---

* Hui Guo, Department of Finance and Real Estate, University of Cincinnati (418 Carl H. Lindner Hall, PO Box 210195, Cincinnati, Ohio 45221-0195, E-mail: hui.guo@uc.edu, and Tel: 513-556-7077). Haim Kassa, Department of Finance and Real Estate, University of Cincinnati (413 Carl H. Lindner Hall, PO Box 210195, Cincinnati, Ohio 45221-0195, E-mail: kassaht@mail.uc.edu, and Tel: 513-556-7087). Michael Ferguson, Department of Finance and Real Estate, University of Cincinnati (415 Carl H. Lindner Hall, PO Box 210195, Cincinnati, Ohio 45221-0195, E-mail: michael.ferguson@uc.edu, and Tel: 513-556-7080). We thank an anonymous referee for many insightful and constructive comments which have greatly improved the exposition of the paper. Hank Bessembinder deserves special thanks for his generous comments on an earlier draft, which greatly improved the exposition of the paper. We are grateful to Doina Chichernea and David Manzler for graciously sharing their SAS codes of estimating EGARCH idiosyncratic volatility and for teaching us how to use them. We thank Fangjian Fu for providing his EGARCH idiosyncratic volatility data and thank Darwin Choi, Fangjian Fu, Qianqiu Liu, and Maria Schutte for explaining their estimation methods of EGARCH idiosyncratic volatility. We have benefited from discussions with Elena Asparouhova, Qing Bai, Turan Bali, Shaun Bond, Doina Chichernea, Brian Hatch, Iva Kalcheva, Andrew Karolyi, Qianqiu Liu, David Manzler, Maria Schutte, Steve Slezak, and seminar and conference participants at University of Cincinnati, the 2011 Financial Management Association meetings, the 2011 Southern Finance Association meetings, the 2012 Southwest Finance Association meetings, and the 2012 Eastern Finance Association meetings. We are responsible for any errors.

On the Relation between EGARCH Idiosyncratic Volatility and Expected Stock Returns

Abstract

A spurious positive relation between EGARCH estimates of expected month t idiosyncratic volatility and month t stock returns arises when the month t return is included in estimation of model parameters. We illustrate via simulations that this look-ahead bias is problematic for empirically observed degrees of stock return skewness and typical monthly return time series lengths. Moreover, the empirical idiosyncratic risk-return relation becomes negligible when expected month t idiosyncratic volatility is estimated using returns only up to month t-1.

I. Introduction

A positive tradeoff between systematic risk and return is the cornerstone of standard rational expectations asset pricing models with a representative agent. However, financial economists have long recognized that idiosyncratic risk is potentially an important determinant of expected stock returns because the portfolio held by a typical U.S. household can only loosely be characterized as diversified (see, e.g., Blume and Friend (1975) and Goetzmann and Kumar (2008)). In particular, many authors, e.g., Levy (1978), Merton (1987), and Malkiel and Xu (2002), argue that these investors—with their poorly diversified portfolios—require extra compensation for holding stocks that expose them to greater idiosyncratic volatility. The empirical evidence for this prediction, however, has been elusive. In fact, Ang, Hodrick, Xing, and Zhang (2006, 2009; hereafter AHXZ) report that the cross-sectional relation between lagged realized idiosyncratic risk and returns is negative. In contrast, Fu (2009) has uncovered a strong positive relation between conditional idiosyncratic volatility estimated using EGARCH models and expected stock returns. Fu's (2009) results are theoretically appealing as well. Consistent with Merton's (1987) conjecture, Fu (2009) documents a *positive* relation between market capitalization and conditional stock returns when controlling for EGARCH idiosyncratic volatility in cross-sectional regressions. Because of its potentially significant contribution to the idiosyncratic risk literature, the EGARCH approach has been widely adopted in related empirical studies (e.g., Spiegel and Wang (2005) and Hwang, Liu, Rhee, and Zhang (2010)).

We must alert researchers, however, that they need to be especially careful when implementing the EGARCH idiosyncratic volatility methodology. A common estimation strategy accidentally introduces a look-ahead bias into recursive volatility forecasts by including the month t return in the estimation of EGARCH parameters that are used to construct the expected month t idiosyncratic volatility. We show analytically that the in-sample EGARCH idiosyncratic volatility can have a strong dependence on the contemporaneous stock return in relatively small samples event though the expected month t idiosyncratic volatility depends on volatility data only through $t-1$. In particular, when we include the month t return in the estimation of EGARCH model parameters, the month t EGARCH idiosyncratic volatility has an upward bias when the month t return is large in magnitude.[1] This bias correlates positively with the month t return if the latter is positive, and the correlation is negative if the latter is negative. Significantly, the cross-section of stock returns is positively skewed (e.g., Duffee (1995)), i.e., there are more stocks with extreme positive returns than stocks with extreme negative returns. Thus, the positive *intertemporal* correlation between the bias in in-sample EGARCH idiosyncratic volatility and one-period-ahead stock returns dominates in stock return data. As a result, the look-ahead bias may generate a spurious predictability of cross-sectional stock returns.

Of course, just because it exists, that does not necessarily imply that the look-ahead bias is so large that it affects statistical inference. This is especially true considering that the 'look ahead' in this case consists of a single monthly return. Therefore, we conduct Monte Carlo simulations to evaluate the impact of the look-ahead bias. The simulation results show that at skewness levels similar to, or even smaller than, those exhibited by monthly CRSP data the look-ahead bias is significant. In addition, the bias is monotonically increasing in skewness. Moreover, despite the fact that the simulations show that the bias is monotonically decreasing in

[1] This result is quite intuitive. We estimate EGARCH models using the maximum likelihood method. An extreme return in month t leads to a particularly low likelihood for this observation. To improve the likelihood of all observations, the month t conditional volatility should increase.

the length of the return series used in estimation, the bias is still significant for return series equivalent to the entire length of the monthly CRSP stock return series.

Having established analytically and via simulations the existence of a potentially significant look-ahead bias in in-sample EGARCH idiosyncratic volatility estimates, we turn to the empirical relation between idiosyncratic risk and the cross-section of stock returns. In one of the best known papers in this literature, Fu (2009) reports three major findings. First, idiosyncratic risk and returns are positively related in the cross-section. Second, after controlling for idiosyncratic risk, there is a positive size effect. Third, he argues that EGARCH idiosyncratic volatility is a better measure of conditional idiosyncratic volatility than the lagged realized idiosyncratic volatility used in AHXZ (2006). We will show that these results are driven by the look-ahead bias introduced by incorporating the month *t* return into the estimate of the month *t* EGARCH idiosyncratic volatility.

We should point out that although we will show that the positive EGARCH idiosyncratic risk-return relation reflects a look-ahead bias, this does not mean that our paper represents an unqualified confirmation of AHXZ (2006).[2] It may well be that the true relationship between idiosyncratic risk and the cross-section of stock returns is positive, as Fu (2009) maintains (or even zero as implied by traditional asset pricing models); but, this particular evidence is

[2] Bali and Cakici (2008) argue that the AHXZ (2006) result is sensitive to different weighting schemes and the estimation of idiosyncratic risk with daily versus monthly return data. Huang, Liu, Rhee, and Zhang (2010) suggest that it relates to the short-horizon return reversal anomaly. Bali, Cakici, and Whitelaw (2011) find that the negative effect of idiosyncratic volatility is driven by its close relation with the maximum daily return in a month, proxying for demand for lottery-like stocks. Jiang, Xu, and Yao (2009) hypothesize that firms with high price volatility tend to be opaque in their earnings disclosures. Han and Lesmond (2011) argue that microstructure noise factors have substantial effects on the realized variance measure.

unreliable due to the look-ahead bias. Therefore, it is premature to conclude that there is strong evidence of a positive relation between idiosyncratic volatility and returns.

To ensure that our results are directly comparable to those reported in existing studies, we use Fu's (2009) monthly estimates of EGARCH idiosyncratic volatility obtained through his website. As a baseline we replicate the result that idiosyncratic risk is positively related to returns. We then show that this is due to the look-ahead bias in two ways. First, we document a strong positive cross-sectional relation between our proxies for the look-ahead bias (e.g., unexpected changes in in-sample EGARCH idiosyncratic volatility) and expected returns. Moreover, including unexpected changes in in-sample EGARCH idiosyncratic volatility in cross-sectional regressions substantially attenuates the explanatory power of the level of in-sample EGARCH idiosyncratic volatility. This result is especially strong when we use log returns instead of simple returns as the dependent variable because log returns have a smaller skewness than do simple returns and, thus, should reduce the magnitude of the look-ahead bias.

The second way that we show that the positive EGARCH idiosyncratic risk and return relation is due to the look-ahead bias is more direct. We simply replace the in-sample EGARCH idiosyncratic risk estimates with our own truly out-of-sample forecast of EGARCH idiosyncratic volatility. In particular, to obtain month t conditional idiosyncratic volatility, we estimate EGARCH model parameters using stock return data up to month $t-1$. We find that while out-of-sample EGARCH idiosyncratic volatility has strong predictive power for one-month-ahead realized idiosyncratic volatility, it does not forecast cross-sectional stock returns.[3] That is, the

[3] Bali, Scherbina, and Tang (2010) have independently verified our main finding that out-of-sample EGARCH idiosyncratic volatility estimates have negligible predictive power for the cross-section of stock returns. After the first draft of this paper was circulated, Fink, Fink, and He (2012) confirmed the weak relation between out-of-sample EGARCH idiosyncratic volatility and the cross-section of stock returns.

positive relation between idiosyncratic risk and returns goes away when we estimate the risk without the look-ahead bias.

In the CRSP data, small capitalization stocks tend to have higher expected returns than large capitalization stocks. Fu (2009), however, shows that this size effect becomes significantly positive after controlling for EGARCH idiosyncratic volatility in cross-sectional regressions. He highlights this finding as direct support for Merton (1987). As with the idiosyncratic risk and return result, we initially illustrate the nature of the bias in in-sample estimates with his own data. We first replicate the result and then we show that the *positive* effect of market capitalization on expected returns disappears when we control for the look-ahead bias in cross-sectional regressions.[4] When we control for out-of-sample idiosyncratic risk we find the traditional size effect. Namely, size is significantly, negatively related to expected returns. Thus, as with the positive idiosyncratic risk and return relationship, we find that when we employ out-of-sample idiosyncratic risk estimates the reported results disappear.

AHXZ (2009) show that lagged realized idiosyncratic volatility has strong explanatory power for one-month-ahead realized idiosyncratic volatility. Fu (2009) suggests that his findings differ qualitatively from AHXZ (2006) because EGARCH idiosyncratic volatility is a better

[4] In Fu's (2009) data, the difference between months *t* and *t-1* EGARCH idiosyncratic volatilities is a proxy for the look-head bias. Specifically, when including both variables in the cross-sectional regression, we show that the former correlates positively with month *t* stock returns, while the relation is negative for the latter. There is a strong negative correlation of month *t-1* EGARCH idiosyncratic volatility with month *t-1* market capitalization. Therefore, a positive relation between month *t-1* market capitalization and month *t* stock returns is found in conjunction with month *t* EGARCH idiosyncratic volatility because the former serves as an instrumental variable for month *t-1* EGARCH idiosyncratic volatility. As expected, the positive size effect goes away when we control for the (unexpected) change in EGARCH idiosyncratic volatility as a proxy for the look-ahead bias.

measure of conditional idiosyncratic volatility than is lagged realized idiosyncratic volatility, for example, as used in AHXZ (2006). We corroborate the AHXZ (2009) finding by showing that the explanatory power of lagged realized idiosyncratic volatility remains statistically significant after controlling for in-sample EGARCH idiosyncratic volatility. In addition, it remains statistically significant after controlling for our out-of-sample EGARCH idiosyncratic volatility estimates. Therefore, lagged realized idiosyncratic volatility provides important information about one-month-ahead realized idiosyncratic volatility beyond EGARCH idiosyncratic volatility. These results cast doubt on the argument that the difference between Fu (2009) and AHXZ's (2006) findings reflects mainly the fact that EGARCH idiosyncratic volatility is a better measure of conditional idiosyncratic volatility than is lagged realized idiosyncratic volatility.

The remainder of the paper proceeds as follows. In Section II, we discuss the look-ahead bias introduced by including the month t return in the estimation of month t EGARCH model parameters. We illustrate the significance of the bias on inference by conducting Monte Carlo simulations. In Section III, we show that the positive cross-sectional relation between EGARCH idiosyncratic volatility and expected stock returns reflects mainly this look-ahead bias. We then show that truly out-of-sample EGARCH idiosyncratic volatility has negligible explanatory power for the cross-section of stock returns. We also revisit Merton's (1987) conjecture that after controlling for idiosyncratic volatility the size effect is positive. In Section IV we conduct robustness tests. In Section V, we discuss the look-ahead bias in EGARCH idiosyncratic volatility estimated using the full sample. In Section VI, we offer some concluding remarks.

II. Look-ahead Bias in In-Sample EGARCH Idiosyncratic Volatility

A. EGARCH Models

Many authors estimate idiosyncratic risk using the Fama and French (1996) three factors as proxies for systematic risk,

$$(1) \qquad R_{i,t} - r_{f,t} = \alpha_i + \beta_i (R_{m,t} - r_{f,t}) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t} \,,$$

where $R_{i,t}$ is stock *i*' return, $r_{f,t}$ is the risk-free rate, $R_{m,t} - r_{f,t}$, SMB_t, and HML_t are the excess market return, the size premium, and the value premium, respectively, as in the Fama and French (1996) three-factor model. The idiosyncratic return, $\varepsilon_{i,t}$, is assumed to have a serially independent normal distribution

$$(2) \qquad \varepsilon_{i,t} \sim N(0, \sigma_{i,t}^2) \,,$$

and its conditional variance, $\sigma_{i,t}^2$, follows an EGARCH process

$$(3) \qquad \ln \sigma_{i,t}^2 = a_i + \sum_{l=1}^{p} b_{i,l} \ln \sigma_{i,t-l}^2 + \sum_{k=1}^{q} c_{i,k} \left\{ \theta \left(\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}} \right) + \gamma \left[\left| \frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}} \right| - (2/\pi)^{1/2} \right] \right\} .$$

Under these assumptions, the log likelihood of the month *t* return, $R_{i,t}$, is

$$(4) \qquad L(R_{i,t}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_{i,t}^2) - \frac{\varepsilon_{i,t}^2}{2\sigma_{i,t}^2} \,.$$

Researchers commonly use the maximum likelihood (or quasi-maximum likelihood if the error term in equation (1) has a nonnormal distribution) method to estimate EGARCH model parameters. That is, the parameter values in equations (1)-(3) are selected to maximize the sum of the log likelihood of stock returns in a given sample period.

B. The Problem with In-Sample Estimates of EGARCH Idiosyncratic Volatility

Many authors rely on the maximum likelihood method to estimate EGARCH model parameters. They refer to the resulting recursive EGARCH idiosyncratic volatility estimates as out-of-sample forecasts. Because equation (3) shows that the month *t* EGARCH idiosyncratic volatility, $\sigma_{i,t}^2$, depends on its own lags and lagged idiosyncratic returns, it is plausible that these estimates do in fact provide an out-of-sample forecast of the month t EGARCH idiosyncratic volatility. However, it is important to note that, if specified inappropriately, this estimation strategy can actually result in an *in-sample* estimate of EGARCH idiosyncratic volatility. Specifically, if we set the sample period to be from month *1* to month *t*, we will include the month *t* stock return in the calculation of the sum of the log likelihood

$$\sum_{\tau=1}^{t} L(R_{i,\tau}) = -\frac{t}{2}\log(2\pi) - \frac{1}{2}\sum_{\tau=1}^{t}\log(\sigma_{i,\tau}^2) - \sum_{\tau=1}^{t}\frac{\varepsilon_{i,\tau}^2}{2\sigma_{i,\tau}^2}. \tag{5}$$

Typically, the next step would be to estimate the EGARCH model by choosing values of the parameters in equations (1)-(3) to maximize the sum of the log likelihood of returns over the period from month *1* to month *t* in equation (5). The problem with this approach is that, via equation (5), the parameter estimates depend (to an asymptotically vanishing degree) on the month *t* return. In particular, the conditional month *t* EGARCH idiosyncratic volatility, $E(IVOL_t)$, has a look-ahead bias because it depends on EGARCH model parameters that are estimated using the month *t* return

$$\begin{aligned} E(IVOL_t) &= \exp(\ln\sigma_{i,t}^2) \\ &= \exp\left[a_{i,t} + \sum_{l=1}^{p} b_{i,l,t}\ln\sigma_{i,t-l}^2 + \sum_{k=1}^{q} c_{i,k,t}\left\{\theta\left(\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right) + \gamma\left[\left|\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right| - (2/\pi)^{1/2}\right]\right\}\right]. \end{aligned} \tag{6}$$

In equation (6), we use the subscript t on the EGARCH model parameter estimates, $a_{i,t}$, $b_{i,l,t}$, and $c_{i,k,t}$ to highlight their dependence on the month t return. Because of its inclusion of the information from month *t*, the conditional idiosyncratic volatility in equation (6), $E(IVOL_t)$, is actually an in-sample EGARCH idiosyncratic volatility estimate. To obtain one-month-ahead forecasts of EGARCH idiosyncratic volatility that are truly out-of-sample, we must restrict equation (5) to include only the returns up to month *t-1* in the calculation of the sum of the log likelihood,

(7) $$\sum_{\tau=1}^{t-1} L(R_{i,\tau}) = -\frac{t-1}{2}\log(2\pi) - \frac{1}{2}\sum_{\tau=1}^{t-1}\log(\sigma_{i,\tau}^2) - \sum_{\tau=1}^{t-1}\frac{\varepsilon_{i,\tau}^2}{2\sigma_{i,\tau}^2}.$$

Obviously equation (7) is obtained from equation (5) by excluding the month *t* return. Then, as is standard, we can estimate an EGARCH model by searching for values of the parameters in equations (1)-(3) that maximize the sum of the log likelihood in equation (7). We then substitute these parameter estimates into equation (3) to obtain the out-of-sample forecast of month *t* EGARCH idiosyncratic volatility, $E(IVOL\_O_t)$,

(8) $$\begin{aligned} E(IVOL\_O_t) &= \exp(\ln\sigma_{i,t}^2) \\ &= \exp\left[a_{i,t-1} + \sum_{l=1}^{p} b_{i,l,t-1}\ln\sigma_{i,t-l}^2 + \sum_{k=1}^{q} c_{i,k,t-1}\left\{\theta\left(\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right) + \gamma\left[\left|\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right| - (2/\pi)^{1/2}\right]\right\}\right]. \end{aligned}$$

In equation (8), we use the subscript *t-1* on the parameter estimates, $a_{i,t-1}$, $b_{i,l,t-1}$, and $c_{i,k,t-1}$ to emphasize the fact that we obtain them using information available at month *t-1*.

The month *t* in-sample EGARCH idiosyncratic volatility from equation (6) has a look-ahead bias because it depends on the month *t* return. This relation is quite intuitive. Suppose that there is an extreme return in month *t* due to an extreme month *t* idiosyncratic return, $\varepsilon_{i,t}$. As a result, equation (4) shows that, ceteris paribus, the log likelihood of the month *t* return is likely to be particularly low. One way to improve the in-sample fit, as illustrated in equation (4), is to raise the month *t* idiosyncratic volatility, $\sigma_{i,t}^2$, by, for example, increasing the constant term, $a_{i,t}$, in conditional volatility of equation (6).[5] That is, the look-ahead bias in the in-sample month t EGARCH idiosyncratic volatility correlates positively with the magnitude of the month t return.

The magnitude of the look-ahead bias depends on the length of the return series used. For example, if we have a large number of stock return observations, the estimates of the EGARCH model parameters in equations (1)-(3) converge asymptotically to their population values. Intuitively, when we have millions of return observations, an extreme return in month t should have a negligible effect on the sum of the log likelihood in equation (5); therefore, the EGARCH parameter estimates do not have to change much to accommodate this extreme observation. In this case, the look-ahead bias in in-sample EGARCH idiosyncratic volatility converges asymptotically to zero. On the other hand, if the number of observations is relatively small, the look-ahead bias can have a substantial impact on the in-sample estimate of EGARCH idiosyncratic volatility. What qualifies as 'relatively small' is an empirical issue that we will investigate in Monte Carlo simulations in sub-section II.D and in an examination of Fu's (2009) major findings in section III.

At this point, the reader may be persuaded that the distortion from including month t returns imparts noise to our volatility estimates but still question whether there is any bias. To see that there often will be, note that when the stock return, $R_{i,t}^{+}$, is positive, it correlates positively with the bias of its in-sample EGARCH idiosyncratic volatility, $EIVOL_{i,t}^{B}$,

(9) $$R_{i,t}^{+} = \alpha EIVOL_{i,t}^{B},$$

[5] Consider equation (4). When conditional idiosyncratic volatility increases, the second term on the right-hand-side (RHS) implies that the log likelihood decreases with log(σ) while the third RHS term implies that the log likelihood increases with σ^2. The more extreme the return, the more the latter dominates the former; hence, the log likelihood increases with conditional idiosyncratic volatility.

where α is a positive parameter. Likewise, when the stock return, $R^-_{j,t}$, is negative, it correlates negatively with the bias of its in-sample EGARCH idiosyncratic volatility, $EIVOL^B_{j,t}$,

(10) $R^-_{j,t} = -\alpha EIVOL^B_{j,t}$.

If cross-sectional stock returns are symmetrically distributed, then the positive relation in equation (9) and the negative relation in equation (10) should approximately cancel each other out in the cross-sectional regression and our idiosyncratic volatility estimates will be noisy but unbiased. However, untabulated results show that cross-sectional stock returns are not symmetrically distributed; but, rather have a strong positive realized skewness in most months, indicating that there are substantially more stocks with extreme positive returns than stocks with extreme negative returns (see also Duffee (1995)). Thus, for the empirical distribution, the positive relation in equation (9) will dominate the negative relation in equation (10) in cross-sectional regressions. That is, there will be a look-ahead *bias*; a positive cross-sectional relation between in-sample EGARCH idiosyncratic volatility and expected stock returns.[6]

In sub-section II.D, we illustrate the dependence of the EGARCH parameter estimates in equation (6) on the month *t* return via Monte Carlo simulations. They verify the existence of a look-ahead bias that is due to a dependence of the month *t* in-sample EGARCH idiosyncratic

[6] The fact that the EGARCH methodology could not truly be making *ex ante* predictions is illustrated by the results of a preliminary experiment of forecasting market returns. Specifically, when we aggregate equations (9) and (10) across all stocks, the equal-weighted market return should correlate positively with the average look-ahead bias if stock returns are positively skewed. As conjectured, we find that changes in average monthly EGARCH idiosyncratic volatility, a proxy for the look-ahead bias, forecast one-month-ahead market returns. A simple switching strategy between a market index and a risk-free Treasury bond based on the information content of EGARCH idiosyncratic volatility generates a Sharp ratio twice as high as buying-and-holding the market index.

volatility on the month *t* return. As the foregoing discussion suggests, the bias is increasing in the skewness of the cross-section of returns and is decreasing in the length of the return series.

C. Estimating Out-of-Sample EGARCH Idiosyncratic Volatility

We use SAS to construct the one-month-ahead out-of-sample forecast of month *t* EGARCH idiosyncratic volatility. That is, we set the sample over the period from month *1* to month *t-1* when estimating EGARCH model parameters. We then substitute the parameter estimates into equation (8) to calculate the month *t* EGARCH idiosyncratic volatility.[7] As in Spiegel and Wang (2005), we require at least sixty monthly return observations to estimate EGARCH models. In contrast, other authors, e.g., Fu (2009), utilizes only thirty monthly return observations. We adopt Spiegel and Wang's (2005) specification because many authors, e.g., Scruggs (1998) and Lundblad (2007), emphasize the need for a large number of observations to obtain precise parameter estimates of GARCH-type nonlinear models. We do not impose a larger minimum because we want our estimates to be comparable to Fu (2009). To further alleviate concerns about the small sample bias, we use an expanding sample starting from July

[7] In an alternative approach, we set the sample over the period from month *1* to month *t* and arbitrarily set the month *t* return to be a missing observation. This effectively 'tricks' SAS into computing an 'in-sample' month *t* EGARCH idiosyncratic volatility that does not depend on information from month *t*. That is, because we have intentionally set the month *t* return to be a missing observation, the EGARCH parameter estimates in equation (6) depend on returns up to month *t-1* and thus have no look-ahead bias. We have confirmed that the EGARCH idiosyncratic volatilities obtained from these two approaches are identical. David Manzler deserves special thanks for suggesting this alternative approach to us.

1926 in the recursive estimations.[8] As in Fu (2009), we consider nine EGARCH specifications, i.e., EGARCH (p,q), where $1 \leq p \leq 3$ and $1 \leq q \leq 3$, and choose the one that converges with the lowest Akaike Information Criterion.

The minimum requirement of sixty monthly return observations is at best a partial solution to the small sample problem. For example, the EGARCH (3,3) model has over ten parameters, and we would not expect to obtain a sensible estimation of these parameters using only sixty observations. Due to the small sample sizes involved, the EGARCH estimates in these studies can be quite sensitive to tuning parameters such as the initial parameter values, the number of iterations, and the convergence criteria. These technical issues highlight the potentially serious problems associated with using EGARCH idiosyncratic volatility estimates. The potential sensitivity of our results raises the possibility that we are accidentally generating the look-ahead bias that we detect below because we are using different convergence criteria, etc. We address this concern in three ways. First, we rely on Fu's (2009) estimates to illustrate the look-ahead bias. Second, when generating our own estimates we select the tuning parameters in such a way that our in-sample EGARCH results closely match those reported in Fu (2009). We then use the same tuning parameters to estimate out-of-sample EGARCH idiosyncratic volatility. In this way, we ensure that the different results obtained from in-sample and out-of-sample EGARCH idiosyncratic volatility estimates reflect only the look-ahead bias. Third, in Section IV, we estimate out-of-sample EGARCH idiosyncratic volatility using a two-year rolling window of daily return data with a minimum of 252 daily returns. The EGARCH estimation is less sensitive to tuning parameters for daily return data because they allow for substantially more

[8] While CRSP monthly stock return data begin in January 1926, the Fama and French three factor data, which we use as a proxy for systematic risk, are available from July 1926.

observations. Again, we find that the EGARCH idiosyncratic volatility has negligible predictive power for the cross-section of stock returns.

D. *Look-ahead Bias: Monte Carlo Simulations*

We have clearly demonstrated that incorporating the month *t* return into the EGARCH estimates induces a bias. On the other hand, we are talking about *one* observation. It is not at all obvious that one observation could have such a statistically meaningful impact on our inferences to account for the positive EGARCH idiosyncratic risk and return relation reported in the previous studies. In this sub-section we examine this very point. We employ Monte Carlo simulation to gauge the effect of the look-ahead bias on statistical inference about the relation between EGARCH idiosyncratic volatility and future stock returns. For simplicity, we generate simulated monthly return data using the EGARCH (1,1) specification. We set the unconditional volatility to 0.06, which is smaller than the median stock return volatility of 0.14 for CRSP common stocks with at least 60 monthly return observations. This conservative calibration ensures that our results are not driven mainly by extreme returns. We set the conditional mean return to zero. Therefore, the simulated data can be interpreted as idiosyncratic returns and by construction, there is no relation between EGARCH idiosyncratic volatility and future stock returns. We generate the i.i.d. error term of the EGARCH model using Ramberg and Schmeiser's (1974) Generalized Lambda Distribution (GLD) algorithm.[9] Specifically, we set the kurtosis of the simulated error term to 3.2, the median kurtosis of CRSP stock returns. In the benchmark case, we set the skewness to 1.1, the median skewness of CRSP common stocks. In simulated data, each stock has 260 return observations, which is slightly more than the median of

[9] We also used Fleishman's (1978) power transformation and found similar results.

230 monthly stock returns for CRSP common stocks. For illustration, we investigate cross-sectional implications using 120 stocks; more stocks should not affect our results in any qualitative manner. Overall, the benchmark case is a reasonably good proxy for the actual data.

As we noted in sub-section II.B, the look-ahead bias increases with skewness. Therefore, we will first attempt to get a feel for the degree of cross-sectional skewness required to affect our inferences by considering different parameterizations of the skewness. Table 1 reports the Fama and MacBeth (1973) estimation results of regressing stock returns on conditional EGARCH volatility obtained using simulated data. For comparison, we consider three specifications. First, under the column "Out of Sample", we estimate the time *t* EGARCH volatility recursively using the information available up to time *t-1*. Second, under the column "In Sample", we estimate the time *t* EGARCH volatility recursively using the information available up to time *t*. Last, under the column "Full Sample", we estimate the EGARCH volatility using the full 260 month sample. In the first two cases, we require a minimum of 60 observations for EGARCH model estimations. Because the simulated data were generated with the EGARCH (1,1) model, we use the same specification to estimate the conditional volatility of the simulated data. We set the maximum iterations to 1,000 and adopt the default SAS convergence tolerance criterion of 0.001. In the cross-sectional regression, we include only stocks that converged in the EGARCH estimations.

[Insert Table 1 here]

Panel D of Table 1 is the benchmark case, in which we set the skewness equal to 1.1. The in-sample EGARCH idiosyncratic volatility correlates positively with future stock returns, and the relation is statistically significant at the 1% level.[10] Note that the presence of a relatively high degree of skewness can facilitate the spurious correlation between expected idiosyncratic volatility and returns. For example, the relation is significantly positive when the skewness is 0.8 (Panel C). For comparison the mean (median) cross-sectional skewness of CRSP common stock returns is 0.8 (1.1). Similarly, we find a positive and significant relation between the full-sample EGARCH idiosyncratic volatility and future stock returns. Moreover, as conjectured, Table 1 shows that the look-ahead bias increases monotonically with the skewness for both the in-sample and full-sample estimates. The out-of-sample results stand in sharp contrast. There is a negligible relation between out-of-sample EGARCH idiosyncratic volatility and future stock returns even for skewness as high as 1.6 (Panel E). These results indicate that, in fact, it is possible for the look-ahead bias to fully account for the positive EGARCH idiosyncratic risk and return relation reported in the literature.[11]

[Insert Table 2 here]

It is a common practice to estimate specifications of the EGARCH model and choose the one that fits the data best according to the Akaike Information Criterion. This specification-selection approach generates an even stronger look-ahead bias than the approach we adopted in Table 1 by using a fixed EGARCH(1,1) specification. Intuitively, when there is an extreme return at time t, the specification that produces the largest time t conditional volatility is most likely to be selected as the best model because it will generally have the highest likelihood. In light of this possibility, we re-run the cross-sectional regression using nine specifications

[10] Interestingly, the coefficient on the EGARCH idiosyncratic volatility in the simulation is 0.118, which is comparable in magnitude to 0.138 obtained using Fu's (2009) data, as reported below in row 1 of Table 4.

[11] For example, Fu (2009) reports a skewness of 2.35 for his data of pooled CRSP common stocks with at least 30 monthly return observations.

(mimicking Fu's (2009) approach) and report the results in Table 2. As expected, the look-ahead bias in Table 2 is noticeably larger than its counterpart in Table 1. For example, in the benchmark case (Panel D), the coefficient on the in-sample EGARCH idiosyncratic volatility is 0.153, compared with 0.118 reported in Table 1. Moreover, a significant positive relation is present at even lower degrees of skewness; a positive relation is present for both the in-sample and full-sample estimates when skewness is 0.4. This is only half (about 1/3rd) the mean (median) cross-sectional skewness of monthly CRSP stock returns.

[Insert Table 3 here]

Lastly, in Table 3, we investigate the effect of the sample size on the look-ahead bias. We consider only the case of the full-sample EGARCH estimation because computation becomes forbiddingly intensive for the recursive in-sample EGARCH estimation when the sample size is large. We set the skewness equal to 0.8 and choose the best model from the nine EGARCH specifications. As conjectured, the look-ahead bias decreases monotonically with the length of the return series; however, it remains statistically significant even when the sample size grows to 5,000 observations. Considering that the entire history of CRSP is roughly 1000 months, this indicates that in practice the monthly return series is never long enough to eliminate the look-ahead bias. For example, in Table 3, when T=1000, the coefficient on the full-sample idiosyncratic volatility is 0.132 with a t-statistic of 6.556. For shorter series, the bias is substantially larger.

Note that, because the look-ahead bias is likely to decrease with the sample size, as we confirm in Table 3, the look-ahead bias of the full-sample EGARCH estimation should be weaker than that of the in-sample EGARCH estimation. In practice, however, we have reason to suspect that full-sample EGARCH idiosyncratic volatility estimates may be prone to even larger *measured* look-ahead biases than in-sample estimates. For a given length return series, say 260 months, the in-sample estimate requires an initial estimation period, say 60 months; therefore there are only 200 months available for the cross-sectional regressions. The full-sample estimates for the same 260 month return series has 60 (or 30%) more months in the cross-sectional regression. As indicated by the simulation results reported in Table 3, the look-ahead bias should be decreasing with the length of the return series. Therefore, we would expect less look-ahead bias in the full-sample. On the other hand, in-sample EGARCH idiosyncratic volatility is estimated over far fewer months and is, hence, noisier. Thus, when we conduct the second stage regression, there will be a more serious error-in-variables problem for the in-sample estimates than for the full-sample estimates. Therefore, the second stage regression coefficients will be more downward biased (the 'attenuation effect') for the in-sample estimates than for the full-sample estimates. It is not clear which one of these effects should dominate; but, in the Monte Carlo simulations reported in Tables 1 and 2 we found the spuriously (yet, frequently, significantly) correlated E[IVOL] coefficients were consistently greater for the full-sample estimates than for the in-sample estimates. So, it appears that the attenuation effect may offset the shorter return series available for the cross-sectional regressions when using in-sample EGARCH estimates.

III. How Important is Look-ahead Bias in Practice?

In Section II we showed analytically and via Monte Carlo simulation that it is possible for EGARCH estimates of idiosyncratic volatility to contain a significant look-ahead bias. In this section we will show that an example of this in practice is Fu's (2009) study of the relation between idiosyncratic risk and the cross-section of stock returns. He reports three major results.

First, he finds a positive cross-sectional relationship between his measure of expected idiosyncratic volatility and returns. Second, he finds a positive relation between firm size and returns (as predicted in Merton (1987)). Third, he suggests that his EGARCH measure of expected idiosyncratic volatility provides a superior forecast of future realized idiosyncratic volatility than the lagged values of realized idiosyncratic volatility advocated in AHXZ (2006).

We will first use Fu's (2009) own estimates of idiosyncratic risk to replicate the positive idiosyncratic risk-return relation he reports. We will then show that this result is attenuated or even eliminated when we control for proxies for the look-ahead bias. More directly, we will then show that there is no relation between EGARCH idiosyncratic volatility and returns when we eliminate the look-ahead bias from the idiosyncratic risk estimates. Additionally, we will show that the positive size effect disappears when we control for the influence of the look-ahead bias. Finally, we will show that the in-sample EGARCH measure of idiosyncratic volatility is superior to lagged realized volatility is also due to the look-ahead bias.

A. *In-Sample EGARCH Idiosyncratic Volatility and the Cross-Section of Stock Returns*

We begin with monthly estimates of stock-level EGARCH idiosyncratic volatility over the July 1963 to December 2007 period obtained from Fangjian Fu at Singapore Management University. We denote this measure E(IVOL) and will frequently refer to it as an in-sample estimate for the reasons outlined in Section II. In row 1 of Table 4, we replicate Fu's (2009) main finding of a strong positive cross-sectional relation between $E(IVOL_t)$ and expected stock returns in the univariate Fama and MacBeth (1973) cross-sectional regression. The point estimate is 0.138 and the adjusted R^2 is 3%, compared with 0.11 and 3%, respectively, as reported in Fu's (2009) Table 5.

[Insert Table 4 here]

Fu (2009) emphasizes that EGARCH idiosyncratic volatility is a good measure of conditional idiosyncratic volatility because it is quite persistent.[12] This finding suggests that $E(IVOL_{t-1})$ should have explanatory power similar to that of $E(IVOL_t)$ for the cross-section of stock returns. This conjecture has also been proposed in a similar context by AHXZ (2006), who show that two-month lagged realized idiosyncratic volatility has explanatory power for the cross-section of stock returns qualitatively similar to that of one-month lagged realized idiosyncratic volatility. Contrary to this conjecture, Table 4 reports that $E(IVOL_{t-1})$ does not correlate with expected stock returns in the univariate cross-sectional regression, with a *t*-statistic close to zero (row 2). More surprisingly, when we include both $E(IVOL_t)$ and $E(IVOL_{t-1})$ as the explanatory variables in the cross-sectional regression, the effect on expected returns remains significantly positive for the former, while it becomes negative and highly significant for the latter (row 3). Because of the strong correlation between $E(IVOL_t)$ and $E(IVOL_{t-1})$, it is tempting to believe that this result reflects a multicollinearity problem. This interpretation, however, does not account for the fact that the *t*-statistics in row 3 are substantially larger in magnitudes than are their univariate counterparts, as reported in rows 1 and 2, respectively.

The result in row 3 of Table 4 may reflect the fact that (unexpected) changes in EGARCH idiosyncratic volatility (which are shocks and, thus, a proxy for the look-ahead bias) have a strong positive correlation with one-month-ahead stock returns. To account for this possibility, we measure unexpected changes in EGARCH idiosyncratic volatility in two ways.

[12] In row 3 of Table 7 below, we confirm this point by showing that one-month lagged EGARCH idiosyncratic volatility, $E(IVOL_{t-1})$, has strong predictive power for realized idiosyncratic volatility, albeit with an adjusted R^2 about 2/3rds that of $E(IVOL_t)$.

First, we consider the difference between $E(IVOL_t)$ and $E(IVOL_{t-1})$, which we dub $\Delta_1 E(IVOL_t)$. Row 4 of Table 4 confirms our conjecture—the first difference, $\Delta_1 E(IVOL_t)$, has a significantly positive correlation with one-month-ahead stock returns even when we control for $E(IVOL_t)$. Because EGARCH idiosyncratic volatility is quite persistent, we also control for the difference between $E(IVOL_t)$ and $E(IVOL_{t-2})$, which we dub $\Delta_2 E(IVOL_t)$. Row 5 of Table 4 reports that $\Delta_2 E(IVOL_t)$ has strong incremental explanatory power as well. In contrast, the explanatory power of $E(IVOL_t)$ attenuates substantially after we control for its changes. For example, in row 5 of Table 4, the parameter estimate and the *t*-statistic of $E(IVOL_t)$ are 0.065 and 2.641, respectively, which are substantially smaller than are their univariate counterparts, as reported in row 1 of Table 4.

As a second proxy for the look-ahead bias, for each stock, we regress its EGARCH idiosyncratic volatility on the two lags, and use the residual from the time-series regression, $UE(IVOL_t)$, as a measure of unexpected changes in $E(IVOL_t)$. In row 6 of Table 4 we see that $UE(IVOL_t)$ has a strong positive correlation with one-month-ahead stock returns in the cross-sectional regression even when we control for $E(IVOL_t)$. Again, the explanatory power of $E(IVOL_t)$ attenuates substantially, as compared with the univariate regression results reported in row 1 of Table 4.

The analysis of Section II, particularly the Monte Carlo simulations, clearly demonstrated that the look-ahead bias in in-sample EGARCH idiosyncratic volatility increases with the skewness of returns. An interesting way to see this in Fu's (2009) data is to use log returns instead of simple returns as the dependent variable in cross-sectional regressions. Naturally, using log returns will alter the distribution of returns, making it more 'Normal' and, thus, reduce the skewness in the data. Moreover, Asparouhova, Bessembinder, and Kalcheva (2011) show that, unlike simple returns, log returns are not subject to biases resulting from microstructure noises. These estimation results are reported in Panel B of Table 4. As expected, when using log returns, we find that the relation between $E(IVOL_t)$ and one-month-ahead stock returns becomes statistically insignificant in the univariate regression at conventional significance levels (row 7). When we control for unexpected changes in $E(IVOL_t)$ in the cross-sectional regression, the relation becomes even negative, and statistically significant in some cases, as shown in rows 10 to 12. Note that in both Panels A and B, unexpected changes in $E(IVOL_t)$ are always significantly positively correlated with expected stock returns. This result reflects the same underlying phenomenon; returns are positively correlated with $E(IVOL_t)$ estimates because the estimates are contaminated by a look-ahead bias.[13]

B. Out-of-Sample EGARCH Idiosyncratic Volatility and the Cross-Section of Stock Returns

[Insert Table 5 here]

We have shown that Fu's (2009) $E(IVOL_t)$ estimates are not positively associated with expected returns when we control for a variety of measures that proxy for the look-ahead bias. We now turn to the most direct demonstration that Fu's (2009) result is due to the look-ahead bias. Table 5 reports on the relation between out-of-sample EGARCH idiosyncratic volatility and expected stock returns. In Panel A, we report the cross-sectional regression results for the July 1963 to December 2006 period, mirroring the sample period in Fu (2009). We also consider

[13] As a robustness check, Fu (2009) also uses log returns in cross-sectional regressions but only tabulates the results for multivariate regressions that include both $E(IVOL_t)$ and market capitalization as independent variables. As we explain in sub-section III.C, this specification strengthens the look-ahead bias and generates a positive size effect.

his three empirical specifications: univariate regression (row 1); controlling for market capitalization and the book-to-market equity ratio (row 2); and, also controlling for past returns, the turnover, and the coefficient of variation of the turnover (row 3). As in Fu (2009), we use the log transformations of firm characteristics except for past stock returns. We find that out-of-sample EGARCH idiosyncratic volatility, $E(IVOL\_O_t)$, has a positive correlation with expected stock returns in all three specifications; however, the correlation is always statistically insignificant at conventional significance levels. Simply put, the positive relation between idiosyncratic risk and returns goes away when we estimate the risk without the look-ahead bias.

As a robustness check, we also consider two different samples, both beginning in September 1931. Because book equity data are unavailable for this early period, we cannot control for the book-to-market equity ratio in cross-sectional regressions for these two samples. Panel B of Table 5 reports the results for the early sample spanning the September 1931 to June 1963 period. We again find a positive, albeit insignificant, relation between $E(IVOL\_O_t)$ and expected stock returns. Second, Panel C reports the results for the full sample spanning the September 1931 to December 2009 period. Row 7 shows that the relation is significantly positive at the 5% level in the univariate regression. It, however, becomes statistically insignificant when we control for market capitalization (row 8). Because small stocks have higher expected returns than do big stocks partly because the former are less liquid (Amihud and Mendelson (1980)), our findings are consistent with those reported by Spiegel and Wang (2005), who document a strong positive relation between EGARCH idiosyncratic volatility and various measures of illiquidity. Row 9 shows that $E(IVOL\_O_t)$ remains statistically insignificant when we control for other commonly used determinants of expected stock returns.

C. *The Positive Size Effect Revisited*

In the CRSP data, there is a pervasive negative relation between market capitalization and expected stock returns. That is, small stocks tend to have higher expected returns than do big stocks. Fu (2009), however, shows that the size effect becomes significantly positive after controlling for EGARCH idiosyncratic volatility in cross-sectional regressions. He highlights this finding as direct support for a novel prediction of Merton's (1987) model—the relation between market capitalization and expected stock returns should be positive when we control for the effect of conditional idiosyncratic volatility on expected stock returns.

In Table 5 we control for out-of-sample idiosyncratic risk and find that size (column 1) is significantly, negatively related to returns in the 1963-2006 period that Fu (2009) analyzes. Furthermore, this holds as well in the 1931-1963 and 1931-2009 periods. Thus, as with the positive idiosyncratic risk and return relationship, we find that when we employ out-of-sample idiosyncratic risk estimates the results disappear.

[Insert Table 6 here]

To further illustrate the nature of the bias in in-sample estimates we again use $E(IVOL_t)$. In Table 6, we first confirm that there is a positive size effect after controlling for $E(IVOL_t)$. Following Fu (2009), we consider two specifications. First, in row 1, we include market capitalization, the book-to-market equity ratio, and $E(IVOL_t)$ as the explanatory variables. Second, in row 3, we add the stock return over the past six months, the turnover, and the coefficient of variation of the turnover to the cross-sectional regression. For both specifications, we replicate the finding of a significantly positive relation between market capitalization and expected stock returns when controlling for $E(IVOL_t)$.

As we did in analyzing the relation between idiosyncratic risk and return in Table 4, we re-run the regressions while controlling for the look ahead-bias. To this end, we include both $\Delta_1 E(IVOL_t)$ and $\Delta_2 E(IVOL_t)$ as proxies for the look-ahead bias, and find that the positive effect of market capitalization on expected stock returns disappears for both specifications (rows 2 and 4).[14] The results reported in row 5 show that controlling for market beta does not qualitatively change our results. In Panel B, we show that the results are qualitatively similar when using log returns as the dependent variable. Our results are quite intuitive. In row 3 of Table 4, we show that while $E(IVOL_t)$ correlates positively with future stock returns, the relation is negative for $E(IVOL_{t-1})$. Because of their strong negative correlation, market capitalization serves as an instrumental variable for $E(IVOL_{t-1})$ when deployed in conjunction with $E(IVOL_t)$. Therefore, its predictive power disappears when we control for the look-ahead bias.

D. Forecasting One-Month-ahead Realized Idiosyncratic Volatility

Fu (2009) suggests that his findings differ qualitatively from those in earlier studies, e.g., AHXZ (2006), because EGARCH idiosyncratic volatility is a better measure of conditional idiosyncratic volatility than is lagged realized idiosyncratic volatility, for example, as used in AHXZ (2006). We investigate this conjecture in Table 7. As in AHXZ (2009), we use the Fama and MacBeth (1973) cross-sectional regression method to investigate the relation between EGARCH idiosyncratic volatility and one-month-ahead realized idiosyncratic volatility.[15] This

[14] We find qualitatively similar results using $UE(IVOL_t)$ as a proxy for the look-ahead bias; for brevity, we do not report these results here but they are available upon request.

[15] Realized volatility is a proxy for a latent variable—the 'true' one-month-ahead volatility—and is estimated with measurement error that is sensitive to methodology, estimation frequency, and estimation window, etc. Therefore,

approach is appropriate because our purpose is to understand the cross-sectional relation between conditional idiosyncratic volatility and stock returns.

[Insert Table 7 here]

As illustrated in Section II, in-sample EGARCH idiosyncratic volatility tends to be high when the one-month-ahead return takes an extreme value, which in turn implies a high realized idiosyncratic volatility. Therefore, the look-ahead bias tends to strengthen the positive relation between $E(IVOL_t)$ and one-month-ahead realized idiosyncratic volatility. To illustrate this point, we report univariate cross-sectional regression results using both in-sample estimate, $E(IVOL_t)$, and out-of-sample estimate, $E(IVOL\_O_t)$, in rows 1 and 2, respectively, of Table 7. While both variables have a strong positive correlation with one-month-ahead realized idiosyncratic volatility, the adjusted R^2 is substantially larger for the former. With this caveat in mind, we discuss below the relative predictive power of EGARCH idiosyncratic volatility versus lagged realized idiosyncratic volatility for one-month-ahead realized idiosyncratic volatility.

In Table 7, we show that, consistent with the results reported in AHXZ (2009), lagged realized idiosyncratic volatility, $IVOL_{t-1}$, has significant predictive power for one-month-ahead realized idiosyncratic volatility (row 4). Noticeably, the adjusted R^2 is 45%, which is substantially higher than the adjusted R^2 of 28% for $E(IVOL_t)$, as reported in row 1. Moreover, row 5 shows that when we include both variables in the cross-sectional regression, lagged realized idiosyncratic volatility remains highly significant, and the adjusted R^2 increases only

---

as a robustness check, we employ future options-implied volatility rather than future realized idiosyncratic volatility as the benchmark. Options-implied volatility is the market's estimate of future volatility. Moreover, it should contain relatively little measurement error because it is available mainly for large optionable stocks. We find qualitatively similar results (untabulated) using the future options-implied volatility as the benchmark.

moderately from 45% in the univariate regression (row 4) to 50% in the multivariate regression. We find qualitatively similar results when controlling for other firm characteristics, including market capitalization, ME, the book-to-market equity ratio, BE/ME, the return over the past six months, RET(-2,-7), the turnover, TURN, and the coefficient of variation of the turnover, CVTURN, in the cross-sectional regression (row 10). Therefore, lagged realized idiosyncratic volatility provides important information about one-month-ahead realized idiosyncratic volatility beyond EGARCH idiosyncratic volatility. These results cast doubt on the argument that the difference between Fu (2009) and AHXZ's (2006) findings reflects mainly the fact that EGARCH idiosyncratic volatility is a better measure of conditional idiosyncratic volatility than is lagged realized idiosyncratic volatility.

IV. Additional Robustness Tests

A. Daily Data

[Insert Table 8 here]

For monthly data, we have to use a relatively small number of return observations to estimate EGARCH idiosyncratic volatility. To address the concern that the EGARCH estimation can be quite sensitive to tuning parameters due to the small sample sizes involved, we estimate the EGARCH idiosyncratic volatility using a two-year rolling window of daily returns with a minimum of 252 observations over the period July 1964 to December 2009.[16] Specifically, we estimate the nine EGARCH specifications using a two-year rolling window of daily stock returns through the last business day of month *t*, and use the specification with the lowest Akaike Information Criterion to make an out-of-sample idiosyncratic volatility forecast for (1) the next day or (2) the next *d* days, where *d* is the number of trading days in month *t+1*. For the first

[16] We thank an anonymous referee for suggesting this alternative measure of EGARCH idiosyncratic volatility.

measure, we multiply the daily idiosyncratic volatility estimate by $\sqrt{22}$ to obtain the expected idiosyncratic volatility of month *t+1*, $E(IVOL\_D1)$. For the second measure, we aggregate *d* conditional daily volatility estimates to get a monthly measure, $E(IVOL\_D2)$. While both alternative EGARCH idiosyncratic volatilities have highly significant predictive power for realized idiosyncratic volatility or options-implied volatility (untabulated), in Table 8, we again find that neither measure forecasts the cross-section of stock returns in either univariate or multivariate regressions. These results cast further doubt on the existing evidence of a positive relation between EGARCH idiosyncratic volatility and future stock returns.

B. Size, Liquidity, and Price Screens

Bali, Cakici, Yan, and Zhang (2005) and Bali and Cakici (2008) show that AHXZ's (2006) finding of a negative relation between realized idiosyncratic volatility and future stock returns is sensitive to a screen for size, price, and illiquidity. As a robustness check, following Bali and Cakici (2008), we exclude (1) the smallest decile stocks by NYSE breakpoints, (2) the most illiquid decile stocks, and (3) stocks with a price below \$10. After screening for size, price, and illiquidity, we sort stocks into two portfolios by market capitalization. Interestingly, we find that in-sample EGARCH idiosyncratic volatility measure forecasts returns only for small stocks but has negligible predictive power for large stocks. This result is not too surprising in light of the Monte Carlo simulation results reported in sub-section II.D above. Small stocks tend to have larger skewness than do large stocks and, thus, are more susceptible to the look-ahead bias in in-sample EGARCH idiosyncratic volatility measure. In contrast, the out-of-sample EGARCH idiosyncratic volatility estimated using either monthly or daily return data always has negligible predictive power for both small and large stocks. For brevity, we do not tabulate these results but they are available on request.

C. *Illiquidity and Idiosyncratic Skewness*

We have shown via simulations that the look-ahead bias in in-sample EGARCH idiosyncratic volatility increases monotonically with skewness. Consistent with this prediction, in Panel B of Table 6, we show that the predictive power of in-sample EGARCH idiosyncratic volatility attenuates substantially when we use log returns. Recent studies, e.g., Boyer, Mitton, and Vorkink (2010), document a strong negative relation between idiosyncratic skewness and future stock returns. Moreover, Bali and Cakici (2008) provide strong evidence for the interaction of illiquidity and idiosyncratic volatility and the effect of this interaction on future stock returns. As a robustness check, we include both idiosyncratic skewness and the Amihud (2002) illiquidity measures as additional control variables and redo the empirical analyses reported in Tables 5 to 7. We find qualitatively similar results (untabulated).

D. *Instrumental Variables*

In Table 5, we reported that out-of-sample EGARCH idiosyncratic volatility has a positive, albeit weak, correlation with expected stock returns. On the other hand, we reported in Table 4 that the correlation with expected returns is significantly positive for in-sample EGARCH idiosyncratic volatility. Some may argue that the latter is a better measure of conditional idiosyncratic volatility than is the former because we need a large number of return observations to obtain precise estimates of EGARCH model parameters. Although investors cannot exploit its correlation with expected stock returns for their portfolio choices, in-sample

EGARCH idiosyncratic volatility is nevertheless useful because it provides a powerful test of economic theories such as Merton's (1987) under-diversification hypothesis. We tested this idea formally using several instrumental variables specifications (available upon request). None of the instrumental variable specifications generated a significantly positive relation between idiosyncratic volatility and expected stock returns.

V. EGARCH Idiosyncratic Volatility Estimated Using the Full Sample

While Fu (2009) estimates EGARCH idiosyncratic volatility recursively, he indicates in his footnote 10 that he finds the same results using the full period data to estimate EGARCH model parameters. Because full-sample EGARCH estimation is computationally less intensive than is recursive EGARCH estimation, many authors, e.g., Brockman and Schutte (2007) and Peterson and Smedema (2011), have subsequently relied upon only the full-sample EGARCH estimates in their studies. The simulation results reported in sub-section II.D suggest that full-sample estimates are subject to a similar look-ahead bias. Therefore, it is not surprising that (in results that, for brevity, are not reported here) we find that the full-sample EGARCH estimates also have the look-ahead bias we documented for Fu's (2009) recursively estimated EGARCH estimates. As we noted in sub-section II.D, full-sample estimates may actually yield even larger cross-sectional coefficients in the second stage regression than recursive in-sample estimates.

VI. Conclusion

We contribute to the empirical literature on the relation between expected idiosyncratic volatility and the cross-section of stock returns by reconsidering findings that EGARCH idiosyncratic volatility is positively related to returns in the cross-section. We show both analytically and empirically that the positive idiosyncratic risk-return relation is driven by a look-ahead bias accidentally introduced by standard methods of estimating month *t* EGARCH idiosyncratic volatility. We show that when month *t* EGARCH idiosyncratic volatility is forecasted using returns only up through month *t-1*, there is no significant cross-sectional relation between EGARCH idiosyncratic volatility and returns. EGARCH estimates can be quite sensitive to the tuning parameters in small samples. To allay fears that this may account for our conclusions, we document that our results continue to hold when we estimate EGARCH volatility from large samples of daily returns.

More generally, we demonstrate that, somewhat counter intuitively, the look-ahead bias introduced by incorporating one extra monthly return is so large that it affects statistical inference. To aid our intuition, we conduct Monte Carlo simulations to evaluate the impact of the look-ahead bias. The simulation results show that at skewness levels similar to, or even smaller than, those exhibited by monthly CRSP data the look-ahead bias is significant. In addition, the bias is monotonically increasing in skewness. Therefore, the bias will be more pronounced in samples that exhibit greater return skewness (e.g., notably, small stocks). Moreover, despite the fact that the simulations show that the bias is monotonically decreasing in the length of the return series used in estimation, the bias is still significant for return series equivalent to the entire length of the monthly CRSP stock return series.

References

Amihud, Y. "Illiquidity and stock returns." *Journal of Financial Markets*, 5 (2002), 31-56.

Amihud, Y., and H. Mendelson. "Dealership Market: Market-Making with Inventory." *Journal of Financial Economics*, 8 (1980), 31-53.

Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259-299.

Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. "High Idiosyncratic Volatility and Low Returns: International and Further U.S. evidence." *Journal of Financial Economics,* 91 (2009), 1-23.

Asparouhova, E.; H. Bessembinder; and I. Kalcheva. "Noisy Prices and Inference Regarding Returns." *Journal of Finance*, forthcoming (2012).

Bali, T., and N. Cakici. "Idiosyncratic Volatility and the Cross-Section of Expected Returns." *Journal of Financial and Quantitative Analysis,* 43 (2008), 29-58.

Bali, T.; N. Cakici; X. Yan; and Z. Zhang. "Does Idiosyncratic Volatility Really Matter?" *Journal of Finance,* 60 (2005), 905-929.

Bali, T.; N. Cakici; and R. Whitelaw. "Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns." *Journal of Financial Economics*, 99 (2011), 427-446.

Bali, T.; A. Scherbina; and Y. Tang. "Unusual News Events and the Cross-Section of Stock Returns." Working Paper, Baruch College (2010).

Blume, M., and I. Friend. "The Asset Structure of Individual Portfolios and Some Implications of Utility Functions." *Journal of Finance,* 30 (1975), 585-603.

Boyer, B.; T. Mitton; and K. Vorkink. “Expected Idiosyncratic Skewness.” *Review of Financial Studies*, 23 (2010), 169-202.

Brockman, P., and M. Schutte. “Is Idiosyncratic Volatility Priced? The International Evidence.” Working Paper, University of Missouri–Columbia (2007).

Duffee, G. “Stock Returns and Volatility: A Firm-Level Analysis.” *Journal of Financial Economics,* 37 (1995), 399-420.

Fama, E., and J. MacBeth. “Risk, Return, and Equilibrium: Empirical Tests.” *Journal of Political Economy,* 71 (1973), 607-636.

Fama, E., and K. French. “Multifactor Explanations of Asset Pricing Anomalies.” *Journal of Finance,* 51 (1996), 55-84.

Fink, J.; K. Fink; and H. He. “Idiosyncratic Volatility Measures and Expected Return.” *Financial Management,* forthcoming (2012).

Fleishman, A. I. “A Method for Simulating Non-normal Distributions.” *Psychometrika,* 43 (1978), 521-532.

Fu, F. “Idiosyncratic Risk and the Cross Section of Stock Returns.” *Journal of Financial Economics,* 91 (2009), 24-37.

Goetzmann, W., and A. Kumar. “Equity Portfolio Diversification.” *Review of Finance,* 12 (2008), 433-463.

Han, Y., and D. Lesmond. “Liquidity Biases and the Pricing of Cross-sectional Idiosyncratic Volatility.” *Review of Financial Studies,* 24 (2011), 1590-1629.

Huang, W.; Q. Liu; S. Rhee; and L. Zhang. "Return Reversals, Idiosyncratic Risk and Expected Returns." *Review of Financial Studies,* 23 (2010), 147-168.

Jiang, G.; D. Xu; and T. Yao. "The Information Content of Idiosyncratic Volatility." *Journal of Financial and Quantitative Analysis,* 44 (2009), 1-28.

Levy, H. "Equilibrium in an Imperfect Market: a Constraint on the Number of Securities in the portfolio." *American Economic Review,* 68 (1978), 643-658.

Lundblad, C. "The Risk Return Tradeoff in the Long Run: 1836-2003." *Journal of Financial Economics,* 85 (2007), 123-150.

Malkiel, B., and Y. Xu. "Idiosyncratic Risk and Security Returns." Working Paper, University of Texas at Dallas (2002).

Merton, R. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance,* 42 (1987), 483-510.

Peterson, D., and A. Smedema. "The return impact of realized and expected idiosyncratic volatility." *Journal of Banking and Finance,* 35 (2011), 2547–2558.

Ramberg, J.S., and B. Schmeiser. "An Approximate Method for Generating Asymmetric Random Variables." *Communications of the ACM,* 17 (1974), 78-82.

Scruggs, J. "Resolving the Puzzling Intertemporal Relation between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach." *Journal of Finance,* 53 (1998), 575–603.

Spiegel, M., and X. Wang. "Cross-sectional Variation in Stock Returns: Liquidity and Idiosyncratic Risk." Working Paper, Yale University (2005).

Table 1

The Impact of Skewness on the Look-ahead Bias: Monte Carlo Simulation Using EGARCH (1,1) Estimation

	Out of Sample	In Sample	Full Sample
	Panel A SKEW = 0.0		
Intercept	-0.000	0.000	0.001
	(-0.099)	(0.007)	(0.665)
E(IVOL)	-0.005	-0.007	-0.023
	(-0.134)	(-0.223)	(-0.732)
Adj. R^2	0.003	0.004	0.006
	Panel B SKEW = 0.4		
Intercept	0.001	-0.001	-0.004*
	(0.670)	(-0.510)	(-2.109)
E(IVOL)	-0.021	0.022	0.069*
	(-0.580)	(0.617)	(2.133)
Adj. R^2	0.004	0.003	0.004
	Panel C SKEW = 0.8		
Intercept	0.002	-0.005*	-0.012**
	(1.217)	(-2.581)	(-6.454)
E(IVOL)	-0.039	0.076*	0.216**
	(-1.418)	(2.362)	(6.203)
Adj. R^2	0.001	0.003	0.007
	Panel D SKEW = 1.1		
Intercept	-0.000	-0.007**	-0.021**
	(-0.012)	(-3.774)	(-11.715)
E(IVOL)	-0.005	0.118**	0.365**
	(-0.167)	(3.609)	(10.722)
Adj. R^2	0.002	0.003	0.007
	Panel E SKEW = 1.6		
Intercept	-0.002	-0.007**	-0.024**
	(-1.464)	(-4.331)	(-14.869)
E(IVOL)	0.043	0.135**	0.450**
	(1.452)	(4.294)	(14.057)
Adj. R^2	-0.000	0.001	0.009

Notes: The table reports the OLS results of the cross-sectional regression of returns on EGARCH estimated idiosyncratic volatility in simulated data. For each of the iterations, 120 artificial stock returns are simulated for 260 months. The unconditional volatility is 0.06 and the conditional mean is set equal to 0. For each of these 120 stock return series, we run out-of-sample, in-sample, and full-sample EGARCH (1,1) to estimate idiosyncratic volatility which we denote by E(IVOL). The default SAS convergence criterion of 0.001 is used, and we set the maximum iterations to 1,000. We require 60 months to start the expanding window volatility estimation for the out-of-sample and in-sample estimates. The two samples differ only in the last return. For each panel, we generate the i.i.d. error term of the EGARCH model using the Generalized Lambda Distribution algorithm of Ramberg and Schmeiser (1974) with kurtosis of 3.2 and the stated skewness. Newey-West corrected *t*-statistics are reported in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.

Table 2

The Impact of Skewness on the Look-ahead Bias: Monte Carlo Simulation Using Nine EGARCH Combinations

	Out of Sample	In Sample	Full Sample
	Panel A SKEW = 0.0		
Intercept	-0.000	-0.000	0.002
	(-0.152)	(-0.344)	(1.695)
E(IVOL)	0.005	0.008	-0.036
	(0.272)	(0.495)	(-1.356)
Adj. R^2	0.005	0.006	0.009
	Panel B SKEW = 0.4		
Intercept	0.001	-0.004*	-0.008**
	(0.877)	(-2.593)	(-5.684)
E(IVOL)	-0.019	0.078*	0.138**
	(-0.734)	(2.568)	(5.574)
Adj. R^2	0.002	0.005	0.006
	Panel C SKEW = 0.8		
Intercept	-0.001	-0.007**	-0.025**
	(-0.476)	(-6.653)	(-18.642)
E(IVOL)	0.003	0.117**	0.435**
	(0.138)	(6.693)	(16.679)
Adj. R^2	-0.000	0.003	0.020
	Panel D SKEW = 1.1		
Intercept	0.001	-0.009**	-0.034**
	(0.623)	(-7.008)	(-28.440)
E(IVOL)	-0.018	0.153**	0.612**
	(-0.871)	(6.611)	(26.195)
Adj. R^2	0.001	0.007	0.037
	Panel E SKEW = 1.6		
Intercept	-0.001	-0.009**	-0.036**
	(-1.621)	(-6.312)	(-25.895)
E(IVOL)	0.025	0.186**	0.689**
	(1.496)	(6.440)	(24.401)
Adj. R^2	-0.001	0.010	0.050

Notes: The table reports the OLS results of the cross-sectional regression of returns on EGARCH estimated idiosyncratic volatility in simulated data. For each of the iterations, 120 artificial stock returns are simulated for 260 months. The unconditional volatility is 0.06 and the conditional mean is set equal to 0. For each these return series, we run out-of-sample, in-sample, and full-sample EGARCH (p,q) for $1 \leq$ (p or q) ≤ 3 to estimate its idiosyncratic volatility which we denote by E(IVOL). The default SAS convergence criterion of 0.001 is used for each of the nine EGARCH combinations, and we set the maximum iterations to 1,000. We select the one that converges with the lowest AIC. We require 60 months to start the expanding window volatility estimation for the out-of-sample and in-sample estimates. The two samples differ only in the last return. For each panel, we generate the i.i.d. error term of the EGARCH model using the Generalized Lambda Distribution algorithm of Ramberg and Schmeiser (1974) with kurtosis of 3.2 and the stated skewness. Newey-West corrected *t*-statistics are reported in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.

Table 3

The Impact of the Length of the Estimation Period on the Look-Ahead Bias: Monte Carlo Simulation Using Nine EGARCH Combinations over the Full Sample with SKEW = 0.8

	T = 100	T = 200	T = 300	T = 400
Intercept	-0.032**	-0.027**	-0.023**	-0.018**
	(-17.860)	(-13.489)	(-11.898)	(-9.436)
E(IVOL)	0.599**	0.482**	0.399**	0.321**
	(17.191)	(12.805)	(11.327)	(8.953)
Adj. R^2	0.053	0.022	0.014	0.011
	T = 500	T = 600	T = 700	T = 800
Intercept	-0.016**	-0.012**	-0.010**	-0.009**
	(-9.381)	(-7.801)	(-7.522)	(-7.994)
E(IVOL)	0.274**	0.202**	0.176**	0.157**
	(8.873)	(7.383)	(7.052)	(7.441)
Adj. R^2	0.008	0.006	0.005	0.004
	T = 900	T = 1000	T = 2000	T = 5000
Intercept	-0.008**	-0.008**	-0.005**	-0.002**
	(-6.788)	(-6.842)	(-6.250)	(-3.208)
E(IVOL)	0.135**	0.132**	0.083**	0.045**
	(6.444)	(6.556)	(6.119)	(3.337)
Adj. R^2	0.004	0.004	0.003	0.006

Notes: The table reports the OLS results of the cross-sectional regression of returns on EGARCH estimated volatility in simulated data. The unconditional volatility is 0.06 and the conditional mean is set equal to 0. For each of the iterations, 120 artificial stock returns are simulated over a sample of T months. For each of these 120 stock returns series, we run full-sample EGARCH (p,q) for $1 \leq p \leq 3$ and $1 \leq q \leq 3$ to estimate its idiosyncratic volatility which we denote by E(IVOL). The default SAS convergence criterion of 0.001 is used for each of the nine EGARCH combinations, and we set the maximum iterations to 1,000. We select the one that converges with the lowest AIC. For each simulation, we generate the i.i.d. error term of the EGARCH model using the Generalized Lambda Distribution algorithm of Ramberg and Schmeiser (1974) with skewness of 0.8 and kurtosis of 3.2. Newey-West corrected *t*-statistics are reported in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.

Table 4

In-Sample EGARCH Idiosyncratic Volatility and Expected Stock Returns

	$E(IVOL_t)$	$E(IVOL_{t-1})$	$\Delta_1 E(IVOL_t)$	$\Delta_2 E(IVOL_t)$	$UE(IVOL_t)$	Adj. R^2
			Panel A Simple Returns			
1	0.138**					0.030
	(6.607)					
2		0.000				0.019
		(0.019)				
3	0.211**	-0.125**				0.037
	(11.746)	(-13.040)				
4	0.086**		0.125**			0.037
	(3.685)		(13.040)			
5	0.065**		0.097**	0.078**		0.040
	(2.641)		(11.854)	(9.946)		
6	0.074**				0.228**	0.039
	(2.990)				(11.471)	
			Panel B Log Returns			
7	0.019					0.026
	(0.983)					
8		-0.070**				0.021
		(-4.068)				
9	0.094**	-0.129**				0.033
	(5.881)	(-13.258)				
10	-0.035		0.129**			0.033
	(-1.553)		(13.258)			
11	-0.057*		0.100**	0.083**		0.037
	(-2.365)		(12.376)	(10.580)		
12	-0.050*				0.238**	0.035
	(-2.066)				(12.005)	

Notes: The table reports Fama and MacBeth (1973) cross-sectional regressions of forecasting one-month-ahead stock returns. $E(IVOL_t)$ is EGARCH idiosyncratic volatility that we obtain from Fangjian Fu at Singapore Management University. $E(IVOL_{t-1})$ is one-month lag of $E(IVOL_t)$. $\Delta_1 E(IVOL_t)$ is the difference between $E(IVOL_t)$ and $E(IVOL_{t-1})$. $\Delta_2 E(IVOL_t)$ is the difference between $E(IVOL_t)$ and its two-month lag, $E(IVOL_{t-2})$. $UE(IVOL_t)$ is the residual from the time-series regression of $E(IVOL_t)$ on a constant and its one-month and two-month lags. We report Newey-West corrected *t*-statistics in parentheses. The data span the July 1963 to December 2006 period. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.

Table 5

Out-of-Sample EGARCH Idiosyncratic Volatility and Expected Stock Returns

	$Ln(ME)$	$Ln(BE/ME)$	$RET(-2,-7)$	$Ln(TURN)$	$Ln(CVTURN)$	$E(IVOL\_O_t)$	Adj. R^2
			Panel A July 1963 to December 2006				
1						0.015	0.014
						(0.995)	
2	-0.089*	0.211**				0.006	0.033
	(-2.296)	(3.714)				(0.475)	
3	-0.145**	0.171**	0.702**	-0.059	-0.453**	0.003	0.054
	(-3.693)	(3.224)	(3.987)	(-0.794)	(-6.077)	(0.460)	
			Panel B September 1931 to June 1963				
4						0.028	0.015
						(1.523)	
5	-0.260**					0.002	0.035
	(-3.021)					(0.130)	
6	-0.309**		0.745	-0.121	-0.300**	0.004	0.070
	(-3.912)		(1.760)	(-1.844)	(-2.614)	(0.373)	
			Panel C September 1931 to December 2009				
7						0.024*	0.015
						(2.185)	
8	-0.208**					0.002	0.030
	(-5.010)					(0.247)	
9	-0.264**		0.700**	-0.089	-0.384**	0.004	0.056
	(-6.727)		(3.307)	(-1.808)	(-6.259)	(0.703)	

Notes: The table reports Fama and MacBeth (1973) cross-sectional regressions of forecasting one-month-ahead stock returns. $Ln(ME)$ is log market capitalization. $Ln(BE/ME)$ is log book-to-market equity ratio. $RET(-2,-7)$ is the return over the previous 7^{th} to 2^{nd} months. $Ln(TURN)$ is log turnover. $Ln(CVTURN)$ is log coefficient of variation of the turnover. $E(IVOL\_O_t)$ is out-of-sample EGARCH idiosyncratic volatility. We report Newey-West corrected *t*-statistics in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.

Table 6

Size, In-Sample EGARCH Idiosyncratic Volatility, and the Cross-Section of Stock Returns

	$Beta$	$Ln(ME)$	$Ln(BE/ME)$	$RET(-2,-7)$	$Ln(TURN)$	$Ln(CVTURN)$	$E(IVOL_t)$	$\Delta_1 E(IVOL_t)$	$\Delta_2 E(IVOL_t)$	Adj. R^2
					Panel A Simple Returns					
1		0.204**	0.444**				0.164**			0.045
		(5.504)	(8.379)				(9.002)			
2		0.049	0.308**				0.070**	0.095**	0.076**	0.053
		(1.437)	(6.191)				(3.205)	(12.422)	(10.942)	
3		0.127**	0.392**	0.910**	-0.360**	-0.730**	0.184**			0.065
		(3.453)	(8.151)	(5.503)	(-5.506)	(-9.040)	(11.782)			
4		-0.003	0.281**	0.891**	-0.218**	-0.579**	0.088**	0.089**	0.072**	0.070
		(-0.096)	(6.011)	(5.530)	(-3.489)	(-7.845)	(4.917)	(12.583)	(11.922)	
5	-0.087	-0.011	0.275**	0.907**	-0.208**	-0.585**	0.090**	0.088**	0.071**	0.076
	(-0.534)	(-0.351)	(6.102)	(6.828)	(-4.053)	(-8.120)	(5.455)	(12.787)	(12.168)	
					Panel B Log Returns					
6		0.223**	0.476**				0.049**			0.043
		(5.976)	(8.685)				(2.983)			
7		0.065	0.412**				-0.047*	0.096**	0.079**	0.051
		(1.923)	(8.502)				(-2.221)	(12.832)	(11.548)	
8		0.146**	0.412**	1.044**	-0.406**	-0.698**	0.074**			0.063
		(3.930)	(8.502)	(6.497)	(-5.760)	(-8.370)	(5.468)			
9		0.017	0.301**	1.025**	-0.264**	-0.547**	-0.023	0.088**	0.073**	0.068
		(0.499)	(6.446)	(6.673)	(-3.975)	(-7.244)	(-1.378)	(13.023)	(12.766)	
10	-0.112	0.007	0.295**	1.037**	-0.248**	-0.551**	-0.020	0.087**	0.071**	0.073
	(-0.673)	(0.223)	(6.549)	(6.983)	(-4.581)	(-7.483)	(-1.310)	(13.320)	(13.203)	

Notes: The table reports Fama and MacBeth (1973) cross-sectional regressions of forecasting one-month-ahead stock returns. $Beta$ is the loading on the market risk. $Ln(ME)$ is log market capitalization. $Ln(BE/ME)$ is log book-to-market equity ratio. $RET(-2,-7)$ is the return over the previous 7th to 2nd months.

$Ln(TURN)$ is log turnover. $Ln(CVTURN)$ is log coefficient of variation of the turnover. $E(IVOL_t)$ is EGARCH idiosyncratic volatility that we obtain from Fangjian Fu at Singapore Management University. $\Delta_1 E(IVOL_t)$ is the difference between $E(IVOL_t)$ and its one-month lag, $E(IVOL_{t-1})$. $\Delta_2 E(IVOL_t)$ is the difference between $E(IVOL_t)$ and its two-month lag, $E(IVOL_{t-2})$. We report Newey-West corrected *t*-statistics in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively. The data span the July 1963 to December 2006 period.

Table 7

The Cross-Section of Expected Idiosyncratic Volatility

	$Ln(ME)$	$Ln(BE/ME)$	$RET(\text{-}2,\text{-}7)$	$Ln(TURN)$	$Ln(CVTURN)$	$E(IVOL_t)$	$E(IVOL\_O_t)$	$E(IVOL_{t-1})$	$IVOL_{t-1}$	Adj. R^2
1						0.631**				0.277
						(45.683)				
2							0.507**			0.163
							(40.069)			
3								0.579**		0.234
								(41.139)		
4									0.678**	0.454
									(85.633)	
5							0.172**		0.618**	0.482
							(46.470)		(65.690)	
6						0.294**			0.550**	0.502
						(57.239)			(56.495)	
7						0.445**		0.322**		0.325
						(48.480)		(39.679)		
8								0.211**	0.582**	0.480
								(33.346)	(57.931)	
9						0.250**		0.104**	0.522**	0.508
						(62.876)		(20.691)	(48.539)	
10	-0.823**	-0.343**	-1.241**	0.376**	0.214**	0.130**			0.482**	0.517
	(-20.357)	(-11.912)	(-9.745)	(10.452)	(4.005)	(25.895)			(42.507)	
11	-0.702**	-0.216**	-1.259**	0.212**	0.041	0.204**		0.063**	0.448**	0.535
	(-19.536)	(-8.604)	(-10.627)	(6.341)	(0.849)	(54.381)		(14.096)	(39.265)	

Notes: The table reports the Fama and MacBeth (1973) cross-sectional regressions of forecasting one-month-ahead realized idiosyncratic volatility. $Ln(ME)$ is log market capitalization. $Ln(BE/ME)$ is log book-to-market equity ratio. $RET(\text{-}2,\text{-}7)$ is the return over the previous 7^{th} to 2^{nd} months. $Ln(TURN)$ is log turnover. $Ln(CVTURN)$ is log coefficient of variation of the turnover. $E(IVOL_t)$ is EGARCH idiosyncratic volatility that we obtain from Fangjian Fu at

Singapore Management University. $E(IVOL_{t-1})$ is one-month lag of $E(IVOL_t)$. $E(IVOL\_O_t)$ is out-of-sample EGARCH idiosyncratic volatility. $IVOL_{t-1}$ is one-month lagged realized idiosyncratic volatility. We report Newey-West corrected *t*-statistics in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively. The data span the July 1963 to December 2006 period.

Table 8

Out-of-Sample EGARCH Idiosyncratic Volatility Estimated Using Daily Return Data

	$Ln(ME)$	$Ln(BE/ME)$	$RET(-2,-7)$	$Ln(TURN)$	$Ln(CVTURN)$	$E(IVOL\_D1)$	$E(IVOL\_D2)$	Adj. R^2
Panel A: July 1964 to December 2006								
1						0.017		0.022
						(1.230)		
2	-0.153**	0.157*				-0.003		0.037
	(-3.980)	(2.530)				(-0.300)		
3	-0.232**	0.117*	0.822**	-0.077	-0.005**	-0.001		0.055
	(-5.840)	(2.090)	(5.140)	(-1.030)	(-6.550)	(-0.080)		
4							0.019	0.027
							(1.130)	
5	-0.165**	0.144*					-0.009	0.040
	(-4.810)	(2.420)					(-0.600)	
6	-0.238**	0.112*	0.811**	-0.070	-0.005**		-0.004	0.057
	(-6.600)	(2.050)	(5.110)	(-0.970)	(-6.890)		(-0.340)	
Panel B: July 1964 to December 2009								
7						0.016		0.021
						(1.170)		
8	-0.169**	0.135*				-0.005		0.036
	(-3.600)	(2.220)				(-0.480)		
9	-0.237**	0.092	0.652**	-0.056	-0.004**	-0.003		0.054
	(-5.180)	(1.640)	(2.730)	(-0.710)	(-4.120)	(-0.410)		
10							0.014	0.027
							(0.860)	
11	-0.191**	0.123*					-0.014	0.039
	(-4.120)	(2.110)					(-1.010)	
12	-0.255**	0.087	0.610*	-0.043	-0.004**		-0.012	0.056
	(-5.640)	(1.600)	(2.380)	(-0.560)	(-4.090)		(-0.950)	

Notes: The table reports Fama and MacBeth (1973) cross-sectional regressions of forecasting one-month-ahead stock returns. $Ln(ME)$ is log market capitalization. $Ln(BE/ME)$ is log book-to-market equity ratio. $RET(\text{-}2,\text{-}7)$ is the return over the previous 7th to 2nd months. $Ln(TURN)$ is log turnover. $Ln(CVTURN)$ is log coefficient of variation of the turnover. $E(IVOL\_D1)$ and $E(IVOL\_D2)$ are out-of-sample EGARCH idiosyncratic volatilities estimated using daily return data. We report Newey-West corrected t-statistics in parentheses. Asterisks * or ** indicate significance at the 5% and 1% levels, respectively.