12-2010

On the robustness of the positive relation between expected idiosyncratic volatility and expected return

Fangjian FU
Singapore Management University, fjfu@smu.edu.sg

On the Robustness of the Positive Relation between Expected Idiosyncratic Volatility and Expected Return

Fangjian Fu
Singapore Management University
December 2010

My 2009 JFE paper ["*Idiosyncratic Risk and the Cross-Section of Expected Stock Returns*", Journal of Financial Economics, Vol. 91, pp. 24-37] documents a positive and statistically significant cross-sectional relation between expected idiosyncratic volatility (E(IVOL)) and expected stock return. A recent working paper titled "*On the Relation between EGARCH Idiosyncratic Volatility and Expected Stock Returns*" by Guo, Ferguson, and Kassa of University of Cincinnati (henceforth GFK) suggests that the positive relation is driven by an in-sample approach to estimate E(IVOL). They fail to find a significant relation between return and their E(IVOL) estimated out of sample.

I have carefully examined Guo, Ferguson, and Kassa's SAS code for the estimation of E(IVOL), and found that two estimation settings in their code, one of which limits the maximum number of iterations and the other accepts estimates with a questionable convergence status, lead to potentially unreliable estimates and ultimately, the failure to find the positive relation between return and E(IVOL). Using more reliable settings, I re-estimate E(IVOL) *strictly out of sample*, and confirm a robust and significantly positive relation between return and E(IVOL), just as reported in my JFE paper.

In the context of risk-return tradeoff, empirical investigations focus on the contemporaneous relation between expected idiosyncratic volatility and expected return. Both variables, however, are not directly observable and need to be estimated. In my JFE paper, I use the conditional volatility estimated by EGARCH models as the measure for E(IVOL).

The E(IVOL) implied by EGARCH (*p, q*) is a function of the **past** *p* periods of residual variances and *q* periods of return shocks. Before applying past information to compute E(IVOL), econometricians need to estimate the *parameters* of the EGARCH model. A typical approach, exemplified in French, Schwert, and Stambaugh (1987, "*Expected Stock Returns and Volatility*", Journal of Financial Economics Vol. 19, pp. 3-29), is to use the full sample period data to estimate GARCH model parameters. Consequently, the conditional volatilities are estimated "within the sample". If the underlying GARCH model is stationary, this approach benefits from more data points and yields more efficient estimates. It is a perfectly acceptable way for researchers to investigate the underlying economic relations.[1] An inconvenience of this approach is, however, that it cannot be exploited to forecast E(IVOL) in the real time. An alternative approach that strives to replicate real-time forecasting is to estimate EGARCH parameters using only the data prior to time *t* and then compute E(IVOL) of time *t*.

For my JFE study, I estimate EGARCH model parameters using an expanding window of data – to estimate E(IVOL) for month t , I truncate return observations after month t. GFK suggest that, in my study, the return of the contemporaneous month is still included in the expanding window to estimate the EGARCH model parameters. Setting the last-month return of the expanding window as missing, they estimate E(IVOL) strictly out of sample, and find no significant relation between their E(IVOL) estimates and stock returns. GFK ascribe the differences in the findings to the inclusion of the contemporaneous return in my estimation of EGARCH model parameters.[2]

Puzzled by their claim, I re-estimate E(IVOL) strictly out of sample (i.e., I also exclude the last-month return of the expanding window), and still find a strong positive relation between E(IVOL) and return, as significant as reported in my JFE paper.

Since the inclusion of the last-month return is not the source for the differences in the findings, I return to scrutinize the differences between their SAS code and mine. I find at least two important differences in the settings for the EGARCH estimation: They are (1) different maximum number of iterations, and (2) different criteria of convergence to determine whether or not an estimate is accepted. As I will show subsequently, these two settings, rather than the inclusion of contemporaneous returns, drive the discrepancies between their results and mine.

Now I explain the two differences in more detail. They set the maximum number of iterations to 100 (maxiter=100), and choose the final estimate from a model that has a convergence status of "0 Converged" or "1 Warning" and yields the lowest AIC (They eliminate models with the convergence status of "2 Failed". After all, a model with failed convergence does not produce any estimates). In contrast, I set maxiter=500 and choose the lowest AIC estimate from a model that **MUST** have the status of "0 Converged".

To illustrate the resulting differences, I choose three stocks as examples: Microsoft, IBM, and Eastman Kodak, and report the results of my explorations in the attached tables.

For each stock, I run three estimations using GFK's SAS code. I first set maxiter=100 (as they did), and then increase maxiter to 500 and 1000, respectively. The contemporaneous return is excluded in all three estimations, so it is not a source for any resulting difference. For each estimation (or equivalently,

[1] Here is another example from an influential paper. Fama and French (1992, "*The Cross-Section of Expected Stock Returns*", Journal of Finance Vol. 48, pp.427-465) use the full period data to estimate the market beta for individual stocks, which are later employed to examine the cross-sectional relation between expected return and beta.

[2] There is only one observation difference in the estimation of EGARCH model. For example, to estimate the E(IVOL) of Microsoft for July 2000, I used the return data from April 1986 to July 2000 (172 return observations) to estimate the parameters of EGARCH models. GFK remove the last-month return (and thus use 171 observations) for the estimation of model parameters. Presumably, the resulting differences in the parameter estimates due to one additional observation should become negligible when the expanding window gets longer.

each maxiter setting), I report the number (and the percentage) of final estimates that are derived from each of the nine EGARCH models (For each stock in each month, I separately employ nine EGARCH (*p, q*) models to estimate E(IVOL) and choose the final estimate from the model that converges and yields the lowest AIC. More details of the estimation can be found in Section 2.3 of my JFE paper). I also report the number (and the percentage) of months in which at least one of the nine EGARCH models ends up in the status of "2 Failed". For these months the final estimate of E(IVOL) is chosen from fewer candidate estimates with the status of "0 Converged" or "1 Warning". I also report the percentage of the final estimates that are at least 1% different in value from those obtained under the setting maxiter=500 (which is the benchmark setting in my explorations). Finally, the last column reports the percentage of the final estimates with the status of "0 Converged"; the rest are generated by models with the status of "1 Warning", which GFK accept, but I reject.

The differences in the estimates resulting from different settings of maxiter are rather striking. When maxiter is increased from 100 to 500, for Microsoft, 66.5% of the final E(IVOL) estimates differ by more than 1% in value; for IBM, 44.9% differ; for Eastman Kodak, 32.1% differ.

The column of "Number (%) of months with failed convergence" provides some clues as to why there is such a large difference. Under the setting of maxiter=100, for Microsoft, in only 6 out of the total 254 months do all the nine EGARCH models converge with "0 Converged" or "1 Warning" (and thus the final estimate is picked out of nine estimates by AIC in these 6 months). In the rest 248, or 97.6% of the months, the number of models that fail to converge ranges from 1 to 8 (median=3, mean=3.3). Clearly, the setting of maxiter=100 is too low, resulting in too many failed convergences.

Once I increase maxiter to 500, there is dramatic improvement in the estimation. The status of "2 Failed" only appears in one month and only for one model in that month. For the rest 253 months, all the nine models converge. Moreover, under maxiter=100, most of the final estimates are generated by EGARCH models (1,1) and (3,1), while under maxiter=500, most of the final estimates are generated by high-order models (3, 3), (3, 1), (3, 2), and (2, 3), suggesting multiple-lagged information (more than just the preceding month) is very useful in forecasting the expected idiosyncratic volatility in the next month. If maxiter is set too low, we fail to utilize the useful lagged information due to the poor convergence of higher-order models. At the extreme, the tests may converge to Ang, Hodrick, Xing, and Zhang (2006), which can be understood as using IVOL(t-1) as an approximation of EIVOL(t) and fails to find the positive trade-off between idiosyncratic risk and return.[3]

These contrasting results suggest that estimation of nonlinear EGARCH models is a challenging task, in particular when most of the nine models involve more than 10 parameters to estimate. The poor convergence performance under maxiter = 100 suggests that this setting is inadequate. Setting maxiter

[3] Ang, Hodrick, Xing, and Zhang (2006. "*The Cross-Section of Volatility and Expected Stock Returns",* Journal of Finance, Vol. 61, pp. 259-299) find a negative relation between stock returns and the *lagged* one-month idiosyncratic volatility [IVOL(t-1)]. My JFE paper shows that stock idiosyncratic volatility varies substantially over time and thus the lagged value is a poor measure of the expected value. Further, this suggests that any other model of expected idiosyncratic volatility could be a poor model if it does not capture more past information than that of the lagged one period.

= 500 improves the estimation performance dramatically. Going further to maxiter = 1000, however, provides little marginal improvement, at least for these several sample stocks.

The setting of convergence criteria is another source for the differences in our estimates. As reported in the last column, about 15% of the final E(IVOL) estimates for Microsoft adopted by GFK are produced by EGARCH models with the status of “1 Warning”. According to the SAS log file, the warning message is “*Good Hessian approximation cannot be obtained. Be careful of interpreting the standard errors*.” This suggests problems with the variance-covariance matrix of the estimate. In my estimation, I only admit the estimates with the status “0 Converged” and avoid estimates with such a warning message. If I use maxiter of 500 but apply GFK’s convergence criteria, 40% of the final estimates are generated from models with a warning message.

Similar conclusions are reached based on the findings of the other two stocks – IBM and Eastman Kodak. Since the stock-level expanding-window estimation is extremely costly in time (more than 1,000 computing hours), I am unable to run their code over the full sample and compare the resulting differences at the full scale. Nonetheless, I believe the evidence from these three example stocks is already telling.

Finally but most importantly, I would like to highlight again that I have re-estimated E(IVOL) strictly out of sample (i.e., using returns up to t-1 for the estimation of E(IVOL) at t), using expanding windows with a requirement of more than 30 months of returns, setting maxiter=500, and admitting estimates only with the status of “0 Converged”, for the sample from July 1926 to December 2009. Using these E(IVOL) estimates, I confirm the positive and statistically significant relation between return and E(IVOL) not only in the sample period of July 63 – Dec 06 as used in my JFE paper, but also in the period before July 63, as well as in the full sample period from July 1926 to Dec 2009. The *t*-statistics for the E(IVOL) slopes in the Fama-MacBeth regressions are close to 10, with or without the controls for other determinants. In short, the results and the conclusions of my JFE paper remain the same and as strong.

Addendum

During the preparation of this note, I was made aware of another working paper ["*Idiosyncratic Volatility Measures and Expected Return*"] by Fink, Fink, and He of James Madison University (Henceforth FFH). The authors find a robust and positive relation between return and the contemporaneous realized idiosyncratic volatility computed from daily returns (which I also report in my 2009 JFE paper), but fail to find a significantly positive relation between return and a few of their out-of-sample forecasts of E(IVOL). One of their forecasts is estimated by EGARCH models. Since I do not have their estimation code, I am unable to analyze the exact reasons for their failure as I did above for GFK. I however would like to offer two general comments on the interpretations of their findings.

FFH compute and denote the difference between the realized idiosyncratic volatility and their estimated E(IVOL) as "*unexpected*" idiosyncratic volatility, and find a strong positive relation between this unexpected idiosyncratic volatility and return. They interpret that the positive relation between return and the realized idiosyncratic volatility is driven by the "unexpected" idiosyncratic volatility and there is no reliable relation between return and the expected idiosyncratic volatility. This interpretation overlooks one possibility that their models are not efficient enough to capture the true expected idiosyncratic volatility. As a result, the component of the realized idiosyncratic volatility that should be "expected" under a more efficient model is misclassified as "unexpected", and vice versa. Evidence from my explorations suggests this alternative explanation is highly possible for their findings.

FFH claims (e.g., in the abstract of their paper) that "A look-ahead bias that has been present in recent papers has led to ***false*** conclusions about the relationship between expected idiosyncratic volatility and expected return." I think it is unthoughtful and misleading to assert that results based on in-sample estimates lead to false conclusions. As exemplified in numerous influential studies, in-sample and out-of-sample estimations of econometric models are two complementary approaches. Each has its own pros and cons and no approach strictly dominates the other. If the goal of the research is to examine the underlying economic relation, as in this case, the underlying economic relation between expected idiosyncratic volatility and expected return, the results based on in-sample estimates of E(IVOL) are certainly not false. As I note earlier, in-sample estimation, relative to out-of-sample estimation, can utilize more data points and yield more efficient estimates (thus a potentially more reliable conclusion on the underlying economic relation).

Ideally we hope to find consistent results with both approaches. FFH find inconsistent results and, without investigating the underlying reasons leading to the differences and the reliability of their particular empirical methodology, immediately conclude one set of results lead to false conclusions. This, again, appears to me unthoughtful and misleading. But never mind, my own explorations yield consistent results under both approaches – a robust and statistically significant positive relation between return and E(IVOL), in which E(IVOL) is estimated either in-sample or strictly out-of-sample.

An exploration on the settings of EGARCH estimation

The following tables summarize the out-of-sample expected idiosyncratic volatility estimates based on Guo, Ferguson, and Kassa's SAS code. I choose three stocks, Microsoft, IBM, and Eastman Kodak, as illustrative examples. There are two important estimation settings: (1) the maximum number of iterations (maxiter) and (2) the convergence status of accepted estimates. In their code, Guo and Kassa set "maxiter=100" and accept estimates with convergence status of 0 or 1. The status of "0 Converged" suggests that the estimation converges for the particular model. The status of "1 Warning", according to the SAS log file, suggests that "Good Hessian approximation cannot be obtained. Be careful of interpreting the standard errors." The tables intend to show how different settings would generate different estimates. In addition to maxiter=100, I also increase maxiter to 500 and 1000, respectively. For the estimation of each maxiter setting, I report the fraction of the final estimates selected with the status of "0 Converged" (the rest are with the status of "1 Warning").

The column of "Number (%) of months with failed convergence" shows the number of months that at least one of the nine EGARCH models fail to converge (thus limit the choice of the final estimate from fewer models). For the column of "Fraction of estimates different by > 1%", I use the estimates under maxiter=500 as the benchmark, and report the percentage of the final estimates in the same month under an alternative maxiter that are more than 1% different in value from the benchmark. The column "Fraction of status = "0 Converged"" reports the percentage of the final estimates chosen with the status of "0 Converged". The sample for Microsoft is from April 1986 to December 2009 (Microsoft went to public in March 1986), and the samples for IBM and Eastman Kodak are from July 1927 to December 2009. I require a stock to have more than 30 months of return observations to be eligible for estimation.

Microsoft (254 months)	Number of final estimates generated by EGARCH(*p*,*q*)									Number (%) of months with failed convergence	Fraction of estimates different by > 1%	Fraction of status = "0 Converged"
	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)			
Maxiter=100	78	0	5	23	11	3	99	17	18	248 (97.6%)	66.5%	85.4%
Maxiter=500	6	1	1	27	8	18	66	48	79	1 (0.4%)	Benchmark	59.8%
Maxiter=1000	6	1	1	27	8	18	66	48	79	0	0.0%	59.8%

IBM (959 months)	Number of final estimates generated by EGARCH(*p*,*q*)									Number (%) of months with failed convergence	Fraction of estimates different by > 1%	Fraction of status = "0 Converged"
	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)			
Maxiter=100	50	57	25	50	140	10	156	236	234	725 (75.6%)	44.9%	89.5%
Maxiter=500	33	51	6	8	94	21	92	340	314	0	Benchmark	78.5%
Maxiter=1000	33	51	6	8	94	21	92	340	314	0	0.0%	78.5%

Eastman Kodak (959 months)	Number of final estimates generated by EGARCH(*p*,*q*)									Number (%) of months with failed convergence	Fraction of estimates different by > 1%	Fraction of status = "0 Converged"
	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)			
Maxiter=100	42	5	10	175	18	57	251	264	137	587 (61.2%)	32.1%	97.0%
Maxiter=500	8	8	5	100	57	79	235	247	220	0	Benchmark	87.3%
Maxiter=1000	8	8	5	100	57	79	235	247	220	0	0.0%	87.3%