COLEMAN FUNG

RISK MANAGEMENT

RESEARCH CENTER

University of California
Berkeley

A Comment on "The Cross-Section of Volatility and Expected Returns": The Statistical Significance of FVIX is Driven by a Single Outlier

Robert M. Anderson[*]   Stephen W. Bianchi[†]

Lisa R. Goldberg[‡]

University of California at Berkeley

December 9, 2012[§]

Abstract

Ang, Hodrick, Xing, and Zhang (2006) examine the pricing of aggregate volatility risk and idiosyncratic risk in the US equity market. As part of that study, they propose an ex post factor, $FVIX$, which is intended as a proxy for aggregate volatility risk at a monthly horizon. Their test validating $FVIX$ as a proxy regresses portfolio excess returns on $FVIX$ and other independent variables over the data period 1986-2000. October 1987 is an outlier, in which the independent variable $FVIX$ exhibits a 26σ deviation from its mean over the data period. The inclusion of a large outlier value of an independent variable results in a spurious *reduction* of the standard error in a regression, in this case by more than a factor of 2. We find that the statistical significance of their tests of $FVIX$ as a proxy disappears when October 1987 is removed from the data set.

[*]Department of Economics, 530 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, USA, email: `anderson@econ.berkeley.edu`.

[†]Department of Economics, 530 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, USA, email: `sbianchi@econ.berkeley.edu`.

[‡]Department of Statistics, 367 Evans Hall #3860, University of California at Berkeley, CA 94720-3860, USA, email:`lrg@berkeley.edu`.

In an important and heavily cited paper, Ang, Hodrick, Xing, and Zhang (2006) examine the pricing of aggregate volatility risk and idiosyncratic risk. Among their striking findings are that the volatility of the aggregate market is a priced risk, and that innovations in aggregate volatility carry a statistically significant negative price of risk of approximately -1% per annum. In addition, they analyze the pricing of idiosyncratic volatility (defined relative to the Fama-French (1993) model) and find that stocks with high idiosyncratic volatility have low average returns: they find a "strongly significant difference of -1.06% per month between the average returns of the quintile portfolio with the highest idiosyncratic volatility stocks and the quintile portfolio with the lowest idiosyncratic volatility stocks."

The results related to aggregate volatility risk in Ang, Hodrick, Xing, and Zhang (2006) rely on an ex post factor, $FVIX$, which is intended as a proxy for aggregate volatility risk.[1] The factor, $FVIX$, is a time-varying portfolio of equities that mimics the daily changes in the original Chicago Board Options Exchange Market Volatility Index.[2] Change in VIX, ΔVIX, is a good proxy for innovation in volatility risk at the daily level. However, volatility exhibits substantial mean reversion. At the monthly level, ΔVIX is contaminated by this mean-reversion, making it unsuitable as a measure of *innovation* in volatility risk.

Ang, Hodrick, Xing, and Zhang (2006, Page 269) explain that $FVIX$, is intended to provide a proxy for innovation in market volatility at a monthly horizon:

> The major advantage of using $FVIX$ to measure aggregate volatility risk is that we can construct a good approximation for innovations in market volatility at any frequency. In particular, the factor mimicking aggregate volatility innovations allows us to proxy aggregate volatility risk at the monthly frequency by simply cumulating daily returns over the month on the underlying base assets used to construct the mimicking factor.

For completeness, we review the construction of $FVIX$. Each month, Ang, Hodrick, Xing, and Zhang (2006) regress daily excess returns for each stock in their dataset, which

includes every common stock listed on the NYSE, AMEX and NASDAQ with more than 17 observations in that month, on the daily excess market return MKT and ΔVIX. They use the estimates $\beta_{\Delta VIX}$ to sort stocks into quintiles; in each quintile, they then form a value-weighted portfolio. Ang, Hodrick, Xing, and Zhang (2006, Table I) report that the quintile portfolios have an average $\beta_{\Delta VIX}$ of -2.09, -0.46, 0.03, 0.54 and 2.18, respectively, where the average is computed over all months in the sample period. These values are called *pre-formation betas*. The most striking analysis concerns the properties of the quintile portfolios in the month after they are formed. The mean monthly returns of the first and fifth quintile portfolios are 1.64% and 0.60% in the *subsequent month*, with the difference having a joint test t-statistic of -3.90. They also compute alphas for the difference, relative to CAPM and the Fama-French 3-factor model, obtaining t-statistics of -3.54 and -2.93.

From Ang, Hodrick, Xing, and Zhang (2006, Page 267):

> While the differences in average returns and alphas corresponding to different $\beta_{\Delta VIX}$ loadings are very impressive, we cannot yet claim that these differences are due to systematic volatility risk. We examine the premium for aggregate volatility within the framework of an unconditional factor model. There are two requirements that must hold in order to make a case for a factor risk-based explanation. First, a factor model implies that there should be contemporaneous patterns between factor loadings and average returns. To test a factor model, Black, Jensen, and Scholes (1972), Fama and French (1992), Fama and French (1993), Jagannathan and Wang (1996), and Pástor and Stambaugh (2003), among others, all form portfolios using various pre-formation criteria, but examine post-ranking factor loadings that are computed over the full sample period. *We must show that the portfolios ... also exhibit high loadings with volatility risk over the same period used to compute the alphas.*" [emphasis added]

For month t, $FVIX_t$ is the time-varying portfolio comprising weights on the quintiles formed in month $t-1$ which best matches ΔVIX in the month t. Ang, Hodrick, Xing, and Zhang (2006) propose $FVIX_t$ as a proxy for volatility risk in month t. The test that the portfolios "exhibit high loadings with volatility risk over the same period used to

compute the alphas" is the monthly regression given in their Equation (6):

$$r_t^i = \alpha^i + \beta_{MKT}^i MKT_t + \beta_{SMB}^i SMB_t + \beta_{HML}^i HML_t + \beta_{FVIX}^i FVIX_t + \varepsilon_t^i \qquad (1)$$

where $i = 1, \ldots, 5$ indexes the quintiles, MKT, SMB and HML are the Fama-French market, size and value factors, $FVIX$ is the mimicking aggregate volatility factor, and the various βs are the corresponding factor loadings. The criteria Ang, Hodrick, Xing, and Zhang (2006) have set for themselves require at least that β_{FVIX}^i vary substantially among the quintiles, and be statistically significant. The final column Ang, Hodrick, Xing, and Zhang (2006, Table I) reports factor loadings of -5.06, -2.72, -1.55, 3.62, and 8.07, with robust Newey-West t-statistics of -4.06, -2.64, -2.86, 4.53, and 5.32, which satisfy the criteria. Our current attempt to replicate the results in Ang, Hodrick, Xing, and Zhang (2006, Table I) is in Panel A of Table I.[3]

However, their data period January 1986–December 2000 includes a significant outlier: October 1987. Note that this month is a significant outlier for two of the *independent* variables: it is a -5.5σ outlier for MKT and a 26σ outlier for $FVIX$.[4] Outliers in a dependent variable can change the regression beta and they increase the standard error. By contrast, outliers in an independent variable may or may not change the regression beta, but they induce a spurious *reduction* in the standard error. A simple explanation of this is in Appendix A.

The regression Ang, Hodrick, Xing, and Zhang (2006, Equation (6)), uses 179 monthly observations; the inclusion of a 26σ outlier in $FVIX$ raises the sample standard deviation of $FVIX$ from the other 178 months by a factor of roughly $\sqrt{\frac{676+178}{178}} \sim 2.19$ and thus spuriously lowers the standard error for β_{FVIX}^i by a factor of 2.

We reran Ang, Hodrick, Xing, and Zhang (2006, Table I) including and excluding October 1987; to check for robustness, we use OLS, Newey-West and Eiker-Hubert-White t-statistics. Including October 1987, *all* fifteen β_{FVIX}^i (five quintiles times three test procedures) estimates have t-statistics greater than 1.96 and thus appear statistically significant. Excluding October 1987, only one in fifteen β_{FVIX}^I values, the OLS estimate for Quintile 1, appears statistically significant. Our replication of the results in Ang, Hodrick, Xing, and Zhang (2006, Table I) with October 1987 omitted is in Panel B of Table I. The differences between the means, CAPM alphas and Fama-French alphas in the fifth and first quintiles remain significant when October 1987 is removed; indeed, all three differences increase in magnitude and in statistical significance. When October 1987 is

4

Table 1: Portfolios Sorted by Exposure to Aggregate Volatility Shocks

Rank	Mean	Std Dev	% Mkt Share	Size	CAPM Alpha	FF-3 Alpha	Pre-Formation $\beta_{\Delta VIX}$	Pre-Formation β_{FVIX}	Next Month Post-Formation $\beta_{\Delta VIX}$	Full Sample Post-Formation β_{FVIX}
A. Full Sample										
1	1.71	5.64	9.15	3.66	0.28	0.30	-1.34	-1.45	-0.024	-0.0412
					[1.49]	[1.69]				[-3.41]
2	1.39	4.44	28.73	4.76	0.13	0.08	-0.43	-0.47	-0.027	-0.0285
					[1.39]	[0.95]				[-4.12]
3	1.34	4.33	30.79	4.76	0.09	0.07	0.03	0.04	0.000	-0.0188
					[0.86]	[0.72]				[-2.92]
4	1.19	4.76	24.04	4.76	-0.14	-0.11	0.51	0.55	0.028	0.0319
					[-1.54]	[-1.22]				[6.15]
5	0.62	6.57	7.29	3.70	-0.91	-0.65	1.49	1.62	0.067	0.0733
					[-3.42]	[-3.23]				[4.27]
5-1	-1.09				-1.19	-0.96				0.1145
	[-3.98]				[-3.39]	[-3.11]				[4.68]
B. Without October 1987										
1	1.89	5.11	9.17	3.66	0.34	0.35	-1.34	-1.45	-0.025	-0.0817
					[1.94]	[1.97]				[-1.89]
2	1.54	4.01	28.79	4.76	0.18	0.12	-0.43	-0.47	-0.027	-0.0220
					[1.80]	[1.50]				[-0.90]
3	1.48	3.93	30.83	4.76	0.12	0.10	0.03	0.04	0.000	0.0031
					[1.20]	[1.16]				[0.19]
4	1.31	4.50	24.00	4.76	-0.18	-0.15	0.51	0.55	0.028	0.0355
					[-2.14]	[-1.91]				[1.27]
5	0.75	6.35	7.20	3.70	-0.99	-0.76	1.49	1.63	0.067	0.0684
					[-3.82]	[-4.13]				[1.42]
5-1	-1.14				-1.34	-1.11				0.1501
	[-4.29]				[-4.11]	[-4.03]				[1.85]

Table I: Following Ang, Hodrick, Xing, and Zhang (2006), we form value-weighted quintile portfolios every month by regressing excess individual stock return on ΔVIX, controlling for the MKT factor, using daily data over the previous month. Stocks are sorted into quintiles based on the coefficient $\beta_{\Delta VIX}$ from lowest (quintile 1) to highest (quintile 5). The statistics in the columns labeled Mean and Std. Dev. are measured in monthly percentage terms and apply to total, not excess, simple returns. Size reports the average log market capitalization for firms within the portfolio. The row "$5 - 1$" refers to the difference in monthly returns between portfolio 5 and portfolio 1. The Alpha columns report Jensen's alpha with respect to the CAPM or the Fama and French (1993) three-factor model. The pre-formation betas refer to the value-weighted $\beta_{\Delta VIX}$ or β_{FVIX} averaged across the whole sample. The second to last column reports the $\beta_{\Delta VIX}$ loading computed over the next month with daily data. The column reports the next month $\beta_{\Delta VIX}$ loadings averaged across months. The last column reports ex post β_{FVIX} over the whole sample, where $FVIX$ is the factor mimicking aggregate volatility risk. To correspond with the Fama-French alphas, we compute the ex post betas by running a four-factor regression with the three Fama-French factors together with the factor that mimics aggregate volatility risk, following the regression in Formula (1). Robust Newey and West (1987) t-statistics are reported in square brackets. Panel A is based on the 180-month dataset beginning in February 1986 and ending in January 2001, including October 1987. Panel B uses the same dataset except October 1987 is omitted.

removed, Full Sample Post-Formation β_{FVIX} of the difference between the fifth and first quintiles increases but its Newey-West t-statistic declines below the cutoff for statistical significance.

The decision to include or exclude major market events is a delicate one. Clearly, financial researchers cannot simply ignore crashes when they try to estimate risk and return. However, one must be careful about what tests are applied and how t-statistics are interpreted, especially when extreme events are included in one's data set.

A An Example of the Effect of a Large Independent Variable Outlier on Ordinary Least Squares (OLS) Regression

For illustration, we consider a one-variable OLS regression. Suppose we are asked to analyze 100 independent observations, $(X_1, Y_1), (X_2, Y_2), \ldots, (X_{100}, Y_{100})$ whose distribution we do not know. The linear regression:

$$Y = \alpha + \beta X + \varepsilon$$

yields estimates:

$$\hat{\beta} = \frac{\sum X_n Y_n - \frac{1}{100} \sum X_n \sum Y_n}{\sum X_n^2 - \frac{1}{100} \left(\sum X_n\right)^2}$$

$$\hat{\alpha} = \frac{1}{100} \left(\sum Y_n - \hat{\beta} \sum X_n\right).$$

of β and α, and an estimate:

$$\hat{\sigma}_\beta = \frac{1}{10} \sqrt{\frac{\sum \epsilon_n^2}{\sum X_n^2}}$$

of σ_β, the standard error of β. Suppose, unbeknownst to us, the data were drawn from a standard bivariate normal distribution, so that $\alpha = \beta = 0$; with these true parameters, $\varepsilon_n = Y_n$ is standard normal and the standard error of beta, σ_β is $1/10$. So we expect to see sample estimates like $\hat{\beta} = .15$ frequently. Given the true distribution, it is legitimate to test the null hypothesis that $\beta = 0$ by comparing the resulting t-statistic to the standard

normal distribution. As long as $\hat{\sigma}_\beta$ is reasonably close to σ_β, we will get a t-statistic that will not lead us to reject the (true) null hypothesis that $\beta = 0$.

Suppose the next draw turns out to be a large outlier that happens to be on the regression line:[5] $(X_{101}, Y_{101}) = (25, \hat{\alpha} + 25\hat{\beta})$. Then our estimates $\hat{\beta}$ and $\hat{\alpha}$ are unchanged but the standard error shrinks by a factor of roughly 2.7:

$$
\begin{aligned}
\hat{\sigma}_\beta(\text{new}) &= \frac{1}{\sqrt{101}} \sqrt{\frac{\sum \epsilon_n^2}{\sum X_n^2}} \\
&= \frac{1}{\sqrt{101}} \sqrt{\frac{\sum_{n=1}^{100} \varepsilon_n^2 + 0}{\sum_{n=1}^{n} X_n^2 + 625\sigma_X^2}} \\
&= \sqrt{\frac{100}{101}} \sqrt{\frac{\sum_{n=1}^{100} X_n^2}{\sum_{n=1}^{100} X_n^2 + 625\sigma_X^2}} \frac{1}{\sqrt{100}} \sqrt{\frac{\sum_{n=1}^{100} \varepsilon_n^2}{\sum_{n=1}^{100} X_n^2}} \\
&\approx \sqrt{\frac{100}{101}} \sqrt{\frac{100\sigma_X^2}{725\sigma_X^2}} \frac{1}{\sqrt{100}} \sqrt{\frac{\sum_{n=1}^{100} \varepsilon_n^2}{\sum_{n=1}^{100} X_n^2}} \\
&\approx \frac{\hat{\sigma}_\beta}{2.7}
\end{aligned}
$$

The presence of the outlier invalidates any statistical inference that compares the t-statistic to a standard normal distribution. If we ignore that point, and naively compare the t-statistic to a standard normal distribution, we will erroneously reject the true null hypothesis that $\hat{\beta} = 0$.

B Generating Apparent Statistical Significance From Pure Noise and a Single Outlier

To shed light on the capacity of a single outlier to generate the illusion of statistical significance in an ordinary least squares regression, we randomly scramble the monthly $FVIX$ returns over time, except for the October 1987 outlier, which is left fixed, and rerun the regression in Formula (1). Figure 1 presents the histograms of Newey-West t-statistics of $FVIX$ betas for Quintiles 1 and 5 resulting from 10^6 scrambles. Using the Gaussian cutoff of 1.96 for statistical significance at the 5% level as a cutoff, we find that 55% of the t-statistics appear to be statistically significant for Quintile 1 and 94% appear to be statistically significant for Quintile 5. Although the scrambling means that 179 of

the 180 monthly values of $FVIX$ are pure noise, a naïve interpretation of the t-statistics that assumes the regression residuals follow a standard normal leads, most of the time, to rejection of the true hypothesis that the $FVIX$ betas are zero.

C   Replication of Ang, Hodrick, Xing, and Zhang (2006, Table I)

Our quintile portfolio means, standard deviations, market shares and sizes are close to the corresponding values reported in Ang, Hodrick, Xing, and Zhang (2006, Table I).

Our pre-formation $\beta_{\Delta VIX}$ coefficients for quintiles 1 and 5 are lower in magnitude than the corresponding values in Ang, Hodrick, Xing, and Zhang (2006, Table I) by roughly 40%. We were able to reconcile with Ang, Hodrick, Xing, and Zhang (2006, Table I) by equally weighting the betas of the individual stocks instead of capitalization weighting them, but we are not sure why that is the right thing to do.

Our post-formation β_{FVIX} coefficients are a factor of 100 lower than the corresponding values in Ang, Hodrick, Xing, and Zhang (2006, Table I). Based on an email message we received from Professor Xing, it seems plausible to us that the $FVIX$ returns were divided by 100 in the regression that produced the results in the last column of Ang, Hodrick, Xing, and Zhang (2006, Table I). This would not affect a key issue, which is the statistical significance of the values in the column, which we replicated. However, we are unsure of the impact it might have on other issues. For example the risk premium associated with $FVIX$ in a four factor regression is reported as -.08 percent per month in Ang, Hodrick, Xing, and Zhang (2006, Table V). If the input returns to $FVIX$ were divided by 100 in that analysis as they were in Ang, Hodrick, Xing, and Zhang (2006, Table I), then the value would by -8 percent per month, or
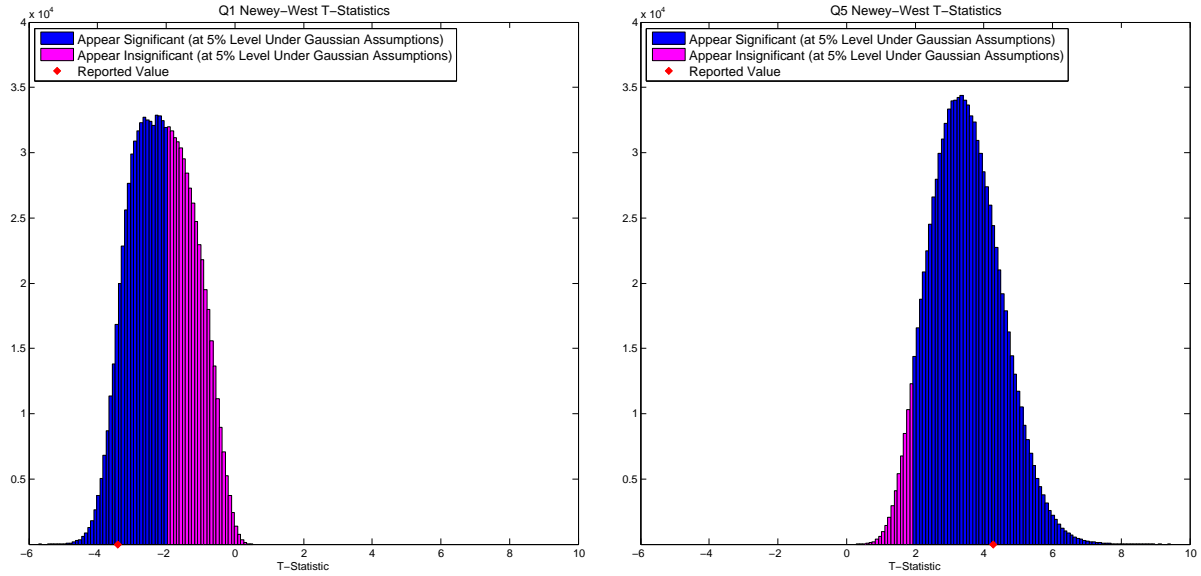
$$100((1 - .08)^{12} - 1)\% = -63.2\%$$

per year.[6]

Figure 1: Histograms of t-statistics for β_{FVIX}. We ran 10^6 repetitions of the four-factor regression in Formula (1) with the returns to $FVIX$ scrambled randomly in time, but with the October 1987 outlier fixed. The histogram of t-statistics for Quintile 1 is in the left panel, and for Quintile 5 is the right panel. Using the Gaussian cutoff of 1.96 for statistical significance at the 5% level as a cutoff, we find that 55% of the t-statistics appear to be statistically significant for Quintile 1 and 94% appear to be statistically significant for Quintile 5.

Notes

[1]The results on idiosyncratic volatility generally do not depend on $FVIX$. An exception is section IIE; Table IX finds that $FVIX$ has limited explanatory power for the findings on idiosyncratic volatility.

[2]The original CBOE Market Volatility Index was launched 1993 under the name VIX and it was based on the Black-Scholes formula. In 2003, the CBOE created a new index based on market prices of call and put options. At that time, they renamed their original index VXO and gave the name VIX to the new index. Ang, Hodrick, Xing, and Zhang (2006) use the index now called VXO, but it is referred to as VIX in their article. To facilitate comparison with the material in their article, we retain the name VIX in this article.

[3]Comments on our replication of Ang, Hodrick, Xing, and Zhang (2006, Table I) are in Appendix C.

[4]To be more precise, the October 1987 $FVIX$ and MKT are 26σ and -5.5σ events relative to the standard deviation of $FVIX$ and MKT over the other 178 months in the sample.

[5]The regression in Equation (1) uses four independent variables: MKT, SMB, HML, and $FVIX$. The quintile returns for October 1987 lie reasonably close (between $6.2°$ and $18.5°$) from the regression hyperplane determined by MKT, SMB and HML, so the case of a large outlier exactly *on* the regression line is reasonably analogous. Because October 1987 is a -5.5σ value of MKT, it carries as much weight in the determination of β_{MKT} as 30 average months. This is the reason the October 1987 quintile returns lie reasonably close to the regression hyperplane.

[6]We have not checked whether this return remains statistically significant once the October 1987 outlier is removed.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* . pp. 79–121 (Praeger Publishers Inc.).

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

———— , 1993, Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *The Journal of Finance* 51, 3–53.

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pástor, Luboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected returns, *The Journal of Political Economy* 111, 642–685.