

In Search of a Statistically Valid Volatility Risk Factor

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

Lisa R. Goldberg[‡]

University of California at Berkeley

February 4, 2013[§]

Abstract

Theory predicts that aggregate volatility ought to be a priced risk factor. In an influential study with more than 1000 citations on Google Scholar, Ang, Hodrick, Xing, and Zhang (2006) propose an ex post factor, *FVIX*, intended as a proxy for aggregate volatility risk. Their test validating *FVIX* relies on an OLS regression of portfolio excess returns on *FVIX* and other independent variables over the data period February 1986–January 2001. October 1987 is an outlier, in which *FVIX* exhibits a 26-sigma deviation. The inclusion of this outlier results in a reduction of the regression standard error by more than a factor of two, creating the appearance of statistical significance when none is present. We explain how standard statistics can be used to assess the suitability of a dataset for OLS regression.

*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.

[§]This research was supported by the Coleman Fung Chair in Risk Management at UC Berkeley. We thank Peter Bickel, Robert Korajczyk, James Sefton, Roger Stein and James Xiong for insightful discussions about the material in this article. We gratefully acknowledge comments from Andrew Ang, Robert J. Hodrick, Yuhang Xing and Xiaoyan Zhang on an early draft, which was entitled “A Comment on ‘The Cross-Section of Volatility and Expected Returns’: The Statistical Significance of FVIX is Driven by a Single Outlier.” Their comments led to substantial improvements in the paper.

Key words: Factor volatility, variance, statistical significance, leverage point, beta, OLS regression, median regression, t -statistic, incremental explanatory power, outlier, VIX , $FVIX$

Since the market exhibits stochastic volatility, equilibrium theory going back to Merton (1973) and Breeden (1979) predicts that aggregate volatility ought to be a priced factor. For example, Anderson and Raimondo (2008) develop an equilibrium model whose primitives are stock dividends, agent endowments, and agent utility functions. If stock dividends or agent endowments (or even agent's utility functions) exhibit stochastic volatility, then except in non-generic "knife-edge" cases, the price of the market portfolio of securities will exhibit stochastic volatility; the volatility of the market portfolio will be a factor which, because it cannot simply be diversified away, generically will have a non-zero price.

There is a large literature exploring the relationship between volatility and asset performance. For example, Carr and Wu (2009) quantify the variance risk premium with portfolios of options. An empirical study by Bollerslev, Tauchen, and Zhou (2009) indicates that a low volatility risk premium predicts high returns and a high volatility risk premium predicts low returns. Sefton, Jessop, Rossi, Jones, and Zhang (2011) attribute the anomalous returns to low-risk investing to variable betas. Adrian and Shin (2010) show that changes in collateralized borrowing and lending on an intermediary's balance sheet are significant forecasting variables for market-wide volatility risk as measured by the VIX Index. In the context of the ICAPM, Campbell, Giglio, Polk, and Turley (2012) develop a vector autoregression that relates aggregate stock market return to volatility shocks. Their model reveals low frequency movements in market volatility that explain the underperformance of growth stocks relative to value stocks. Cremers, Halling, and Weinbaum (2012) use investable option strategies to model aggregate jump and volatility risk in the cross-section of asset returns.

A common feature of the diverse collection of empirical articles exploring the relationship between volatility and asset performance is their reference to Ang, Hodrick, Xing, and Zhang (2006). Among the striking findings in that article, which has more than 1000 citations on Google Scholar, 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. The results rely on an ex post factor, $FVIX$,

which is intended as a proxy for aggregate volatility risk.

In the course of our research, we followed the same path as many others did and turned to *FVIX* in order to gain insight into the nature of volatility risk. In an attempt to make sense of empirical results that we found puzzling, we replicated the construction of *FVIX*. We also replicated the five monthly regressions used in Ang, Hodrick, Xing, and Zhang (2006) to establish *FVIX* as a risk factor. The dependent variable in each regression is a quintile portfolio of assets sorted by sensitivity to daily changes in the VIX index.

We found that when the regression t -statistics are correctly interpreted, *FVIX* betas are not statistically significant in any of the five regressions intended to establish the efficacy of *FVIX*. This is documented in Table I, which shows the impact of removing a single leverage point (outlier of an independent variable), October 1987, from the dataset: the magnitudes of the Newey-West t -statistics of the *FVIX* betas fall below the Gaussian cutoff of 1.96. The point is not that the coefficients change when October 1987 is omitted; they do change, but not dramatically. Rather, the point is that the October 1987 leverage point reduces the reported standard error of β_{FVIX} by more than a factor of two in each of the five regressions, making insignificant coefficients appear to be statistically significant.

The potential for a leverage point or another type of influential observation to render a model unsuitable for estimation with an OLS regression is well known in the statistics literature; see, for example Chatterjee and Hadi (1986). However, standard measures for detecting and managing influential observations are not part of standard practice in empirical finance. The leverage points examined in this article are important from an economic perspective since they represent actual market events. They are also interesting and subtle from a statistical perspective: they materially deflate the volume of regression confidence ellipsoids but have a limited effect on the coefficients themselves, thereby enabling pure noise to masquerade as statistical significance.

Our paper is organized as follows. In Section 1, we review the construction of *FVIX* and the statistics used by Ang, Hodrick, Xing, and Zhang (2006) to support its efficacy. In Section 2, we argue that the OLS regressions used to establish the statistical significance of *FVIX* are inappropriate and lead to erroneous inference. In Section 3, we examine a standard statistic used to measure the impact of a leverage point on statistical significance.

In Section 4, we re-examine the statistical significance of β_{FVIX} with a median regression, which is less sensitive to leverage points and other types of outliers than OLS regression. Section 5 concludes.

We explore some of the statistical properties of the October 1987 leverage points in the first two of three appendices to this article. In Appendix A, we construct a simple example that shows how a leverage point can generate apparent statistical significance when none is present. Specifically we use OLS regression to estimate a one-factor model whose underlying data is Gaussian with the exception of a single leverage point. The example is crafted to highlight some of the interesting properties of the data set used to estimate β_{FVIX} . To further illustrate the pitfalls of neglecting leverage points, we present in Appendix B a simulation showing how to generate apparent statistical significance from pure noise and a single leverage point. While keeping the October 1987 value in place, we randomly scrambled the other $FVIX$ values. The final Appendix documents discrepancies between Ang, Hodrick, Xing, and Zhang (2006, Table I) and our replication of that table.

1 FVIX

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.¹ As emphasized in Ang, Hodrick, Xing, and Zhang (2006), 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 $\beta_{\Delta VIX}$ estimates 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 Ang, Hodrick, Xing, and Zhang (2006, equation (6)):

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

where $i = 1, \dots, 5$ 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 β s are the corresponding factor loadings. The criteria Ang, Hodrick, Xing, and Zhang (2006) have set for themselves require that β_{FVIX}^i be large and vary substantially across 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 replication of the results in Ang, Hodrick, Xing, and Zhang (2006, Table I) is in Panel A of Table I.²

2 The Apparent Statistical Significance of $FVIX$ Betas is Driven by a Single Leverage Point

The data period in Ang, Hodrick, Xing, and Zhang (2006), February 1986–January 2001, includes a leverage point (a significant outlier of an *independent* variable): October 1987. Indeed, this month is a leverage point for two of the independent variables: it is a -5.5-sigma outlier for MKT and a 26-sigma outlier for $FVIX$.³ Outliers in a dependent variable can change the regression beta and increase the standard error. By contrast, leverage points may or may not change the regression beta, but they reduce the standard error. A simple explanation of this is in Appendix A.

The regression defined by Ang, Hodrick, Xing, and Zhang (2006, Equation (6)) uses 180 monthly observations; the inclusion of a 26-sigma leverage point in $FVIX$ raises the sample standard deviation of $FVIX$ from the other 180 months by a factor of roughly $\sqrt{\frac{676+179}{179}} \sim 2.19$ and lowers the OLS standard error for β_{FVIX}^i by a factor of 2. It has an even greater effect on the Newey-West t -statistics used by Ang, Hodrick, Xing, and Zhang (2006). The usual cutoff of $t = \pm 1.96$ depends on the assumption that the data are Gaussian. These data are clearly not Gaussian, and t -statistics exceeding 1.96 are not

sufficient to establish statistical significance. This is illustrated in Appendix B, where we demonstrate that a leverage point can generate the illusion of statistical significance from pure noise.

The results of our replication of Ang, Hodrick, Xing, and Zhang (2006, Table I) with and without October 1987 are in Panels A and B of Table I. Panel A summarizes the analysis on the full dataset; all five quintile *FVIX* beta estimates have *t*-statistics greater than 1.96 and thus appear statistically significant if one misapplies the Gaussian cutoff. Panel B summarizes the analysis on the same dataset with the October 1987 leverage point excluded. None of the five quintile *FVIX* beta estimates are greater than 1.96 in magnitude. The exclusion of October 1987 makes only modest changes in the regression coefficients but results in a dramatic reduction of the *t*-statistics. When the 26-sigma leverage point is excluded, there is no reduction of the standard error of β_{FVIX} . The *t*-statistics for the differences between the means, CAPM alphas and Fama-French alphas in the fifth and first quintiles exceed 1.96 in magnitude when October 1987 is removed; indeed, all three differences increase in magnitude and exhibit higher *t*-statistics. When October 1987 is removed, the Full Sample Post-Formation β_{FVIX} of the difference between the fifth and first quintiles increases but the magnitude of its Newey-West *t*-statistic declines below the 1.96 cutoff.

Panels C and D of Table I summarize an update to the study in Ang, Hodrick, Xing, and Zhang (2006) that adds 131 observations by including data through December 2011. For the most part, the results over the longer period are qualitatively similar to the results over the shorter period, although the additional data dilute the effect of the October 1987 leverage point. On the full dataset, the Newey-West *t*-statistics for *FVIX* declined, in some cases substantially, for Quintiles 1 through 4, although it increased from 4.27 to 5.61 in Quintile 5. When October 1987 is removed, the additional 131 observations raised the magnitudes of some of the *FVIX* Newey-West *t*-statistics and lowered the magnitudes of other, but the magnitudes of all of them remained below the 1.96 cutoff.

Table I: 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}
<i>A. Full Sample (1986-2001)</i>										
1	1.71	5.64	9.15	3.66	0.28 [1.49]	0.30 [1.69]	-1.34	-1.45	-0.024	-0.0412 [-3.41]
2	1.39	4.44	28.73	4.76	0.13 [1.39]	0.08 [0.95]	-0.43	-0.47	-0.027	-0.0285 [-4.12]
3	1.34	4.33	30.79	4.76	0.09 [0.86]	0.07 [0.72]	0.03	0.04	0.000	-0.0188 [-2.92]
4	1.19	4.76	24.04	4.76	-0.14 [-1.54]	-0.11 [-1.22]	0.51	0.55	0.028	0.0319 [6.15]
5	0.62	6.57	7.29	3.70	-0.91 [-3.42]	-0.65 [-3.23]	1.49	1.62	0.067	0.0733 [4.27]
5-1	-1.09 [-3.98]				-1.19 [-3.39]	-0.96 [-3.11]				0.1145 [4.68]
<i>B. Without October 1987 (1986-2001)</i>										
1	1.89	5.11	9.17	3.66	0.34 [1.94]	0.35 [1.97]	-1.34	-1.45	-0.025	-0.0817 [-1.89]
2	1.54	4.01	28.79	4.76	0.18 [1.80]	0.12 [1.50]	-0.43	-0.47	-0.027	-0.0220 [-0.90]
3	1.48	3.93	30.83	4.76	0.12 [1.20]	0.10 [1.16]	0.03	0.04	0.000	0.0031 [0.19]
4	1.31	4.50	24.00	4.76	-0.18 [-2.14]	-0.15 [-1.91]	0.51	0.55	0.028	0.0355 [1.27]
5	0.75	6.35	7.20	3.70	-0.99 [-3.82]	-0.76 [-4.13]	1.49	1.63	0.067	0.0684 [1.42]
5-1	-1.14 [-4.29]				-1.34 [-4.11]	-1.11 [-4.03]				0.1501 [1.85]
<i>C. Full Sample (1986-2011)</i>										
1	1.20	5.96	9.77	4.25	0.27 [1.75]	0.27 [1.83]	-1.25	-0.85	-0.005	-0.0363 [-2.31]
2	0.98	4.54	28.97	5.33	0.18 [2.32]	0.16 [2.34]	-0.40	-0.26	-0.026	-0.0443 [-2.57]
3	0.90	4.55	30.64	5.38	0.09 [1.29]	0.05 [0.95]	0.03	0.06	-0.005	-0.0178 [-2.58]
4	0.83	5.08	23.08	5.28	-0.03 [-0.40]	-0.08 [-1.06]	0.49	0.37	0.021	0.0218 [1.77]
5	0.28	7.15	7.55	4.25	-0.75 [-4.20]	-0.72 [-4.58]	1.42	1.02	0.074	0.0946 [5.61]
5-1	-0.92 [-4.02]				-1.03 [-4.50]	-0.99 [-4.17]				0.1309 [4.47]
<i>D. Without October 1987 (1986-2011)</i>										
1	1.30	5.69	9.78	4.25	0.24 [1.66]	0.21 [1.45]	-1.25	-0.84	-0.005	-0.0371 [-1.19]
2	1.06	4.31	29.00	5.33	0.15 [2.42]	0.15 [2.45]	-0.40	-0.26	-0.026	-0.0336 [-1.72]
3	0.98	4.35	30.66	5.39	0.05 [0.71]	0.05 [0.71]	0.03	0.06	-0.005	0.0104 [0.75]
4	0.89	4.95	23.05	5.28	-0.11 [-1.35]	-0.11 [-1.36]	0.49	0.37	0.021	0.0156 [0.81]
5	0.35	7.04	7.50	4.25	-0.86 [-4.78]	-0.87 [-5.26]	1.42	1.02	0.075	0.0712 [1.52]
5-1	-0.95 [-4.20]				-1.09 [-5.04]	-1.08 [-4.93]				0.1083 [1.65]

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. Panels A and B are based on the 180-month dataset, February 1986–January 2001. Panels C and D are based on the 331-month dataset February 1986–December 2011. Panels A and C include all data; Panels B and D exclude the October 1987 leverage point.

3 A Standard Measure of a Leverage Point’s Influence on Statistical Significance

The volume of the regression confidence ellipsoid has a major effect on confidence testing.⁴ The smaller the ellipsoid, the easier it is to reject a null hypothesis. A standard statistic that measures the influence of an observation on the confidence ellipsoid volume is the covariance ratio, denoted CVR . The statistic is based on the determinant of the covariance matrix of estimated betas, which is inversely related to the square of the confidence ellipsoid volume. By definition, CVR_t is the ratio of the determinants of the covariance matrices of estimated betas with and without observation t :

$$CVR_t = \frac{\det(s_t^2(X_t^\top X_t)^{-1})}{\det(s^2(X^\top X)^{-1})},$$

where X is the matrix of independent variables, s^2 is the sum of the squared residuals and the subscript, t , indicates that the t th observation has been excluded from the dataset. Belsley, Kuh, and Welsch (1980) and Chatterjee and Hadi (1986) suggest that $3p/T$ is a reasonable upper bound for values of $|CVR - 1|$, where p is the number of independent variables (including the intercept) and T is the number of observations.

For each quintile regression, and for the regression of the difference between quintiles 1 and 5, we estimated CVR for each observation. The results for the regression over the longer period that includes data through December 2011 are shown in Figure 1. Observations that are influential according to the $3p/T$ rule are marked in red. October 1987 is a standout in every case. This confirms that the inclusion of October 1987 in the OLS regression artificially lowers the volume of the confidence ellipsoid in a way that makes insignificant findings appear significant.

4 Median Regression

There is a large literature devoted to the analysis of datasets that do not satisfy the assumptions that justify OLS estimation; see, for example, Koenker (2005). One of the most accessible techniques is a *median regression*, which determines coefficients β by minimizing the objective:

$$\sum_t |y_t - x_t\beta - \alpha|. \tag{2}$$

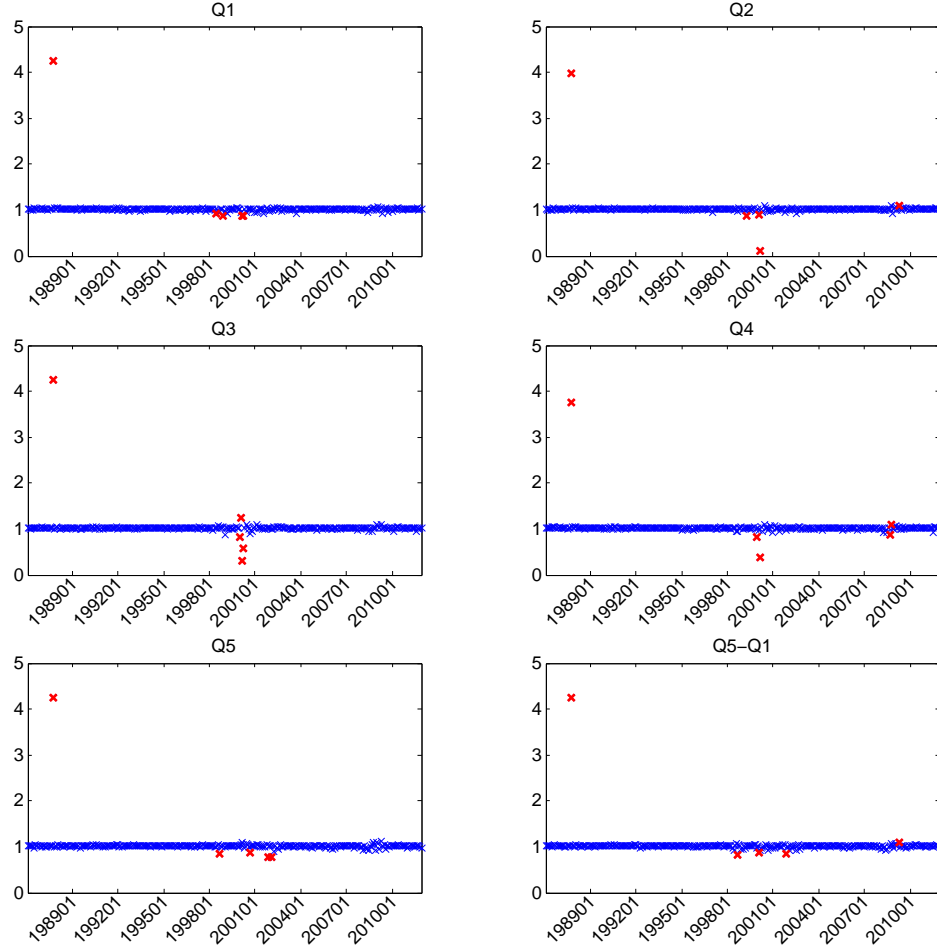


Figure 1: Covariance ratios for observations in an OLS estimation of six four-factor models based on Fama and French (1993) plus *FVIX* over the longer time horizon, February 1986–December 2011. The dependent variables are the five Quintile portfolios and the difference between Quintiles 1 and 5. Outliers satisfying $|CVR - 1| > 3p/T$ are marked in red. The October 1987 leverage point is a standout in each regression.

Since Formula 2 features the absolute values of residuals and not their squares, leverage points and other types of outliers have less impact on the results of median regression than on the results of OLS regression. We estimate betas with a median regression, and we explore the statistical significance of the estimates with bootstrapped percentile confidence intervals. This method for estimating statistical significance is widely used because it is conceptually simple and easily implementable. However, it must be used with caution when the histogram of the bootstrapped betas is unstable, as this suggests that the underlying assumptions of the bootstrap may not be satisfied.

Using the objective in Formula 2, we ran four median regressions based on Formula 1: over the shorter and longer time horizons, and with and without October 1987. The results are displayed in Table II. Over shorter time horizon, February 1986–January 2001, the results indicate that on the full dataset, β_{FVIX} appears to be statistically significant for Quintiles 4 and 5, but not for the other quintiles. When October 1987 is omitted, β_{FVIX} does not appear to be statistically significant in any of the regressions. Over the longer horizon, February 1986–December 2011, β_{FVIX} appears to be statistically significant for Quintiles 2 and 5, with and without October 1987, but not for the other quintiles.

How much credence do we give to these results? To answer that question, we examine the histograms of bootstrapped betas used to determine the percentile confidence intervals. Histograms of 100,000 bootstrapped estimates of β_{FVIX} for Quintile 4 for each of the four median regressions each are displayed in Figure 2. The histograms in the left panel of Figure 2 are based on the full datasets. The inclusion of October 1987 distorts the histograms with a dramatic skew, which is to the right over the shorter time period and to the left over the longer time period. When October 1987 is omitted, the inclusion of the extra 131 observations through December 2011 broaden and flatten the histogram while leaving an interesting spike toward the right. The non-Gaussian shapes, the instability over time, and the dramatic impact of a single observation lead us to suspect that with or without October 1987, the assumptions underlying the bootstrap are not satisfied. Based on this and the mixed nature of the results over quintiles, the median regressions do not support the conclusion that FVIX is a statistically significant volatility factor.

While a median regression appears to be better suited to our dataset than an OLS regression, a more sophisticated median regression that allows for a mixture of distributions may be better still. We continue to pursue this line of inquiry.

Table II: Median Regression β_{FVIX}

	Full Sample		
	Post-Formation	Percentile	
Rank	β_{FVIX}	Confidence Interval	
<i>A. Full Sample (1986-2001)</i>			
1	-0.0307	-0.1361	0.0577
2	-0.0265	-0.0461	0.0130
3	-0.0238	-0.0306	0.0357
4*	0.0305	0.0085	0.0806
5*	0.0640	0.0136	0.1228
5-1	0.0975	-0.0061	0.1998
<i>B. Without October 1987 (1986-2001)</i>			
1	-0.0736	-0.1657	0.0878
2	-0.0241	-0.0584	0.0261
3	-0.0061	-0.0371	0.0427
4	0.0571	-0.0098	0.0944
5	0.0782	-0.0206	0.1491
5-1	0.1400	-0.0518	0.2359
<i>C. Full Sample (1986-2011)</i>			
1	-0.0292	-0.0916	0.0698
2*	-0.0272	-0.0725	-0.0124
3	-0.0195	-0.0274	0.0362
4	0.0329	-0.0357	0.0420
5*	0.0837	0.0396	0.1395
5-1	0.1229	-0.0304	0.1878
<i>D. Without October 1987 (1986-2011)</i>			
1	-0.0202	-0.1203	0.0931
2*	-0.0361	-0.0805	-0.0038
3	-0.0025	-0.0347	0.0433
4	0.0040	-0.0464	0.0486
5*	0.1054	0.0114	0.1563
5-1	0.0953	-0.0601	0.2313

Table II: Percentile confidence intervals for median regressions. Panels A and B are based on the 180-month dataset, February 1986–January 2001. Panels C and D are based on the 331-month dataset February 1986–December 2011. Panels A and C include all data; Panels B and D exclude the October 1987 leverage point. Over the shorter time horizon, February 1986–January 2001, on the full dataset, β_{FVIX} appears to be statistically significant for Quintiles 4 and 5, but not otherwise. When October 1987 is omitted, β_{FVIX} does not appear to be statistically significant in any of the regressions. Over the longer horizon, February 1986–December 2011, β_{FVIX} appears to be statistically significant for Quintiles 2 and 5, with and without October 1987, but not for the other quintiles.

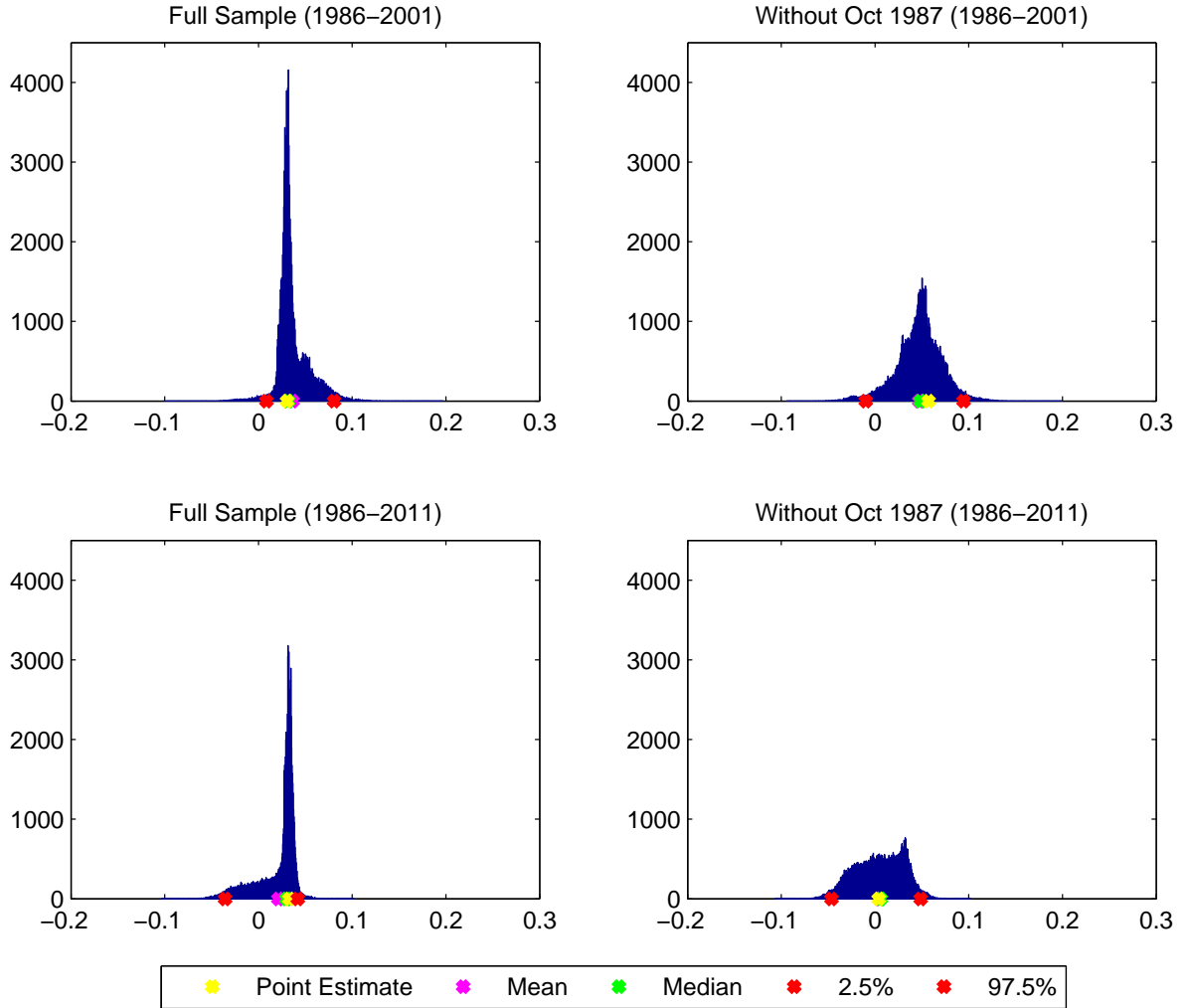


Figure 2: Histograms of bootstrapped values of β_{FVIX} estimated using median regression. The histograms in the left panel include the October 1987 leverage point and the histograms in the right panel omit it. The histograms in the top panels are based on the shorter period, February 1986–January 2001, and the histograms in the bottom panels are based on the longer period, February 1986–December 2011.

5 Conclusion

Ordinary least squares regression is the bedrock of empirical finance. However statistical inference based on OLS regression is valid only when the underlying dataset is Gaussian. The real-world example discussed in this article illustrates how a single leverage point can generate the illusion of statistical significance, leading to an unsound conclusion that has echoed throughout the financial economics literature.

Since extreme events are endemic to financial markets, there is a compelling case for requiring the reporting of standard statistics for detecting influential observations. Standard measures of the impact of influential observations on regression coefficients, fitted values, and t -statistics can enable researchers to avoid erroneous inference and to apply regression techniques that are appropriate for the data they are examining.

There is no single procedure for handling leverage points and other statistical anomalies. Financial researchers cannot simply ignore crashes in the estimation of risk and return; removal and winsorization can mask important characteristics of a dataset. Median regression and other quantile-based estimation techniques, which are less sensitive to the statistical extremes, may be more suitable than OLS regression for many financial data sets. While much work remains to be done in order to identify the tools that are fully appropriate for analyzing financial data, the inclusion of standard statistics as part of best practices in empirical finance can prevent erroneous conclusions from being drawn.

A A Simple Example of How a Leverage Point Can Lead to False Inference in OLS Regression

For illustration, we consider a one-variable model. Suppose we are asked to analyze 100 independent observations, $(X_1, Y_1), (X_2, Y_2), \dots, (X_{100}, Y_{100})$ whose distribution we do not know. If we fit the model:

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

with an OLS regression, we obtain estimates:

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

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

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, $\epsilon_n = Y_n$ 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: $(X_{101}, Y_{101}) = (25, \hat{\alpha} + 25\hat{\beta})$. Then our estimates $\hat{\beta}$ and $\hat{\alpha}$ are unchanged but the standard error shrinks by a factor of roughly 2.7:

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

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 3 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 found we could match the figures in Ang, Hodrick, Xing, and Zhang (2006, Table I) by equally weighting the betas of the individual stocks instead of capitalization weighting them, and conjecture that Ang, Hodrick, Xing, and Zhang (2006) inadvertently used equal weighting.

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, we believe that Ang, Hodrick, Xing, and Zhang (2006) regressed on *FVIX*/100 rather than on *FVIX*; with this understanding, we qualitatively replicate their coefficients. Regressing on *FVIX*/100 rather than on *FVIX* has no effect on the reported *t*-statistics, which we were able to replicate qualitatively. However, it

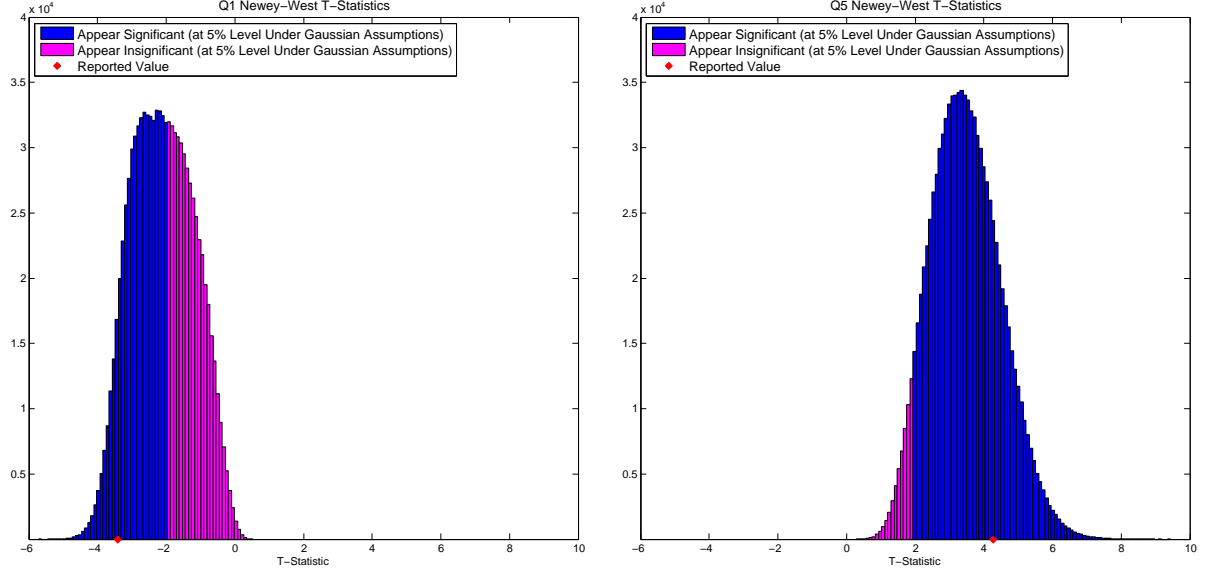


Figure 3: Histograms of t -statistics for β_{FVIX} . We ran 10^6 repetitions of the four-factor regression in Formula (1) with the returns to $FVIX$ scrambled randomly in time, but with the October 1987 outlier fixed. The histogram of t -statistics for Quintile 1 is in the left panel, and for Quintile 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.

increases the reported Full Sample Post-Formation β_{FVIX} numbers by a factor of 100, making it appear that the post-formation β_{FVIX} spreads are much larger than the post-formation $\beta_{\Delta VIX}$ spreads, which Ang, Hodrick, Xing, and Zhang (2006) characterize as “disappointingly very small.” Our estimate of the monthly return to $FVIX$ is approximately 100 times larger than reported by Ang, Hodrick, Xing, and Zhang (2006); we believe they calculated the return to $FVIX/100$.

Given that Ang, Hodrick, Xing, and Zhang (2006) reported that $FVIX$ is a statistically significant factor, we were surprised to find no discussion of its incremental explanatory power over and above the factors in Fama and French (1993). Out of curiosity, we examined the question. A standard F -test shows that on the complete dataset, $FVIX$ appears to have incremental explanatory power in all quintiles. However, when the October 1987 leverage point is removed, $FVIX$ does not add explanatory power for Quintiles 2, 3, 4 and 5.

Notes

¹The original CBOE Market Volatility Index was launched in 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.

²Detailed comments on our replication of Ang, Hodrick, Xing, and Zhang (2006, Table I) are in Appendix C.

³To be more precise, the October 1987 $FVIX$ and MKT are 26-sigma and -5.5-sigma events relative to the standard deviation of $FVIX$ and MKT over the other 179 months in the sample.

⁴A δ -confidence ellipsoid for an OLS regression with p independent variables (including the intercept) and T observations is given by:

$$(\beta - \hat{\beta})^\top \frac{(X^\top X)}{s^2} (\beta - \hat{\beta}) \leq \frac{(T-1)p}{T(T-p)} F_{p, T-p}(\delta).$$

References

Adrian, Tobias, and Hyun Song Shin, 2010, Liquidity and leverage, *J. Finan. Intermediation* 19, 418–437.

- Anderson, Robert M., and Roberto C. Raimondo, 2008, Equilibrium in continuous-time financial markets, *Econometrica* 76, 841–907.
- 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.
- Belsley, David A., Edwin Kuh, and Roy E. Welsch, 1980, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (Wiley).
- 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.).
- Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *The Review of Financial Studies* 22, 4463–4492.
- Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.
- Campbell, John Y., Stefano Giglio, Christopher Polk, and Robert Turley, 2012, An intertemporal capm with stochastic volatility, NBER Working Paper 18411.
- Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *The Review of Financial Studies* 22, 1311–1341.
- Chatterjee, Sampreet, and Ali S. Hadi, 1986, Influential observations, high leverage points, and outliers in linear regression, *Statistical Science* 1, 379–393.
- Cremers, Martijn, Michael Halling, and David Weinbaum, 2012, Aggregate jump and volatility risk in the cross-section of stock returns, Johnson School Research Paper Number 13-2010. Available at SSRN: <http://ssrn.com/abstract=1565586> or <http://dx.doi.org/10.2139/ssrn.1565586>.
- 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.
- Koenker, Roger, 2005, *Quantile Regression* (Cambridge University Press).
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 876–887.
- 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.
- Sefton, James, David Jessop, Giuliano De Rossi, Claire Jones, and Heran Zhang, 2011, Low-risk investing, UBS Investment Research.