Section 7: Volatility

Ralph S.J. Koijen Stijn Van Nieuwerburgh*

April 5, 2024

*Koijen: University of Chicago, Booth School of Business, NBER, and CEPR. Van Nieuwerburgh: Columbia Business School, CEPR, and NBER. If you find typos, or have any comments or suggestions, then please let us know via ralph.koijen@chicagobooth.edu or svnieuwe@gsb.columbia.edu.

1. Basic structure of the notes

- High-level summary of theoretical frameworks to interpret empirical facts.
- Per asset class, we will discuss:
  1. Key empirical facts in terms of prices (unconditional and conditional risk premia) and asset ownership.
  2. Interpret the facts using the theoretical frameworks.
  3. Facts and theories linking financial markets and the real economy.
  4. Active areas of research and some potentially interesting directions for future research.
- The notes cover the following asset classes:
  1. Equities (weeks 1-5).
     - Discount rates and the term structure of risk (week 1)
     - The Cross-section and the factor zoo (week 2)
     - Intermediary-based Asset Pricing (week 3)
     - Production-based asset pricing (week 4)
     - Demand-based asset pricing (week 5)
  2. Mutual funds and hedge funds (week 6).
  3. Volatility (week 7).
  4. Government bonds (week 8).
  5. Corporate bonds and CDS (week 9).
  6. Currencies and international finance (week 10).
  7. Commodities (week 11).
  8. Real estate (week 12).

2. Volatility Facts

2.1. Measuring Variance

- Ways to measure or estimate variance:

  1. Parametric models (e.g., GARCH).
  2. Realized variance using high-frequency data.
  3. Implied volatility using options.
  4. Text-based variance measures.

- Before high-frequency data were available, the conditional variance of returns, $\sigma_t^2 = V_t(R_{t+1})$, was measured via GARCH-style models.

- High-frequency data makes variance effectively observable, but GARCH-style models are still useful for predicting future volatility.

2.1.1. Parametric models

- GARCH(p,q) model (Bollerslev, 1986)

$$
\begin{aligned}
r_{t+1} &= \mu + \eta_{t+1}, \\
\eta_{t+1} &= \sigma_t \varepsilon_{t+1}, \\
\sigma_t^2 &= \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \eta_{t+1-j}^2,
\end{aligned}
$$

  where $\varepsilon_{t+1} \sim N(0,1)$, i.i.d., $\alpha_j \geq 0$, $\beta_i \geq 0$, $\omega \geq 0$.

- To capture the persistence of volatility with the original ARCH(q) model (Engle, 1982), we need many lags.
- The same logic as ARMA processes in time-series econometrics.
- By now, there are lots of extensions of such models, see Bollerslev (2008) for an overview.
- There are also many multivariate extensions modeling the conditional covariance matrix of returns (BEKK, VECH, DCC, . . .).
- GARCH(1,1) is a commonly-used specification

$$\sigma_t^2 = \omega + \alpha \eta_t^2 + \beta \sigma_{t-1}^2.$$

- Note that

$$\sigma_t^2 = \omega + (\alpha + \beta)\, \eta_t^2 - \beta \left(\eta_t^2 - \sigma_{t-1}^2 \right),$$

  and after adding η_{t+1}^2on both sides

$$\eta_{t+1}^2 = \omega + (\alpha + \beta)\, \eta_t^2 - \beta \left(\eta_t^2 - \sigma_{t-1}^2 \right) + \left(\eta_{t+1}^2 - \sigma_t^2 \right),$$

it shows that a GARCH(1,1) can be viewed as an ARMA(1,1) for η_{t+1}^2, where $\eta_{t+1}^2 - \sigma_t^2$ are the innovations.

- Shocks to $\eta_t^2 - \sigma_{t-1}^2$ feed into η_{t+1}^2 in the next period at rate $\beta < 1$ and then persist into the following periods at rate $\alpha + \beta$ (close to 1).

- This means that $\alpha + \beta$ governs the persistence of conditional variance in the GARCH(1,1) model.

- Also, we need $\alpha + \beta < 1$ to ensure covariance stationarity.

- Applying the basic ARMA(1,1) logic to see that the unconditional mean of the conditional variance equals

$$\sigma^2 = E(\sigma_t^2) = E(\eta_t^2) = \frac{\omega}{1 - \alpha - \beta}.$$

- We can use GARCH models for forecasting.

- One-period forecast

$$E_t\left(\eta_{t+1}^2\right) = \sigma_t^2.$$

- Two-period forecast

$$\begin{aligned} E_t\left(\eta_{t+2}^2\right) &= \omega + (\alpha + \beta)\,\sigma_t^2 \\ &= \sigma^2 + (\alpha + \beta)\left(\sigma_t^2 - \sigma^2\right). \end{aligned}$$

- $k-$period forecast

$$E_t\left(\eta_{t+k}^2\right) = \sigma^2 + (\alpha + \beta)^{k-1}\left(\sigma_t^2 - \sigma^2\right).$$

- Estimation is typically done via Maximum Likelihood
- Denote $\theta = (\mu, \alpha, \beta, \omega)$,

$$r_{t+1} \mid \sigma_t^2 = N\left(\mu, \sigma_t^2\right).$$

- The conditional density is

$$l\left(r_{t+1} \mid \sigma_t^2; \theta\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{1}{2}(r_{t+1}-\mu)^2/\sigma_t^2}.$$

- The conditional likelihood of the data

$$\mathcal{L}\left(r_T, ..., r_1 \mid \sigma_0^2; \theta\right) = \prod_{t=1}^{T} l\left(r_t \mid \sigma_{t-1}^2; \theta\right).$$

- We maximize this likelihood over θ.
- The likelihood is conditional on σ_0^2.
  - It is common to set σ_0^2 equal to the unconditional variance.
  - Asymptotically, the impact of the pre-sample values disappears, but it matters in finite samples in particular when $\alpha + \beta$ is close to 1.
- The asymptotic covariance matrix of the ML estimator is calculated (as always) based on the second derivative of the log likelihood w.r.t. the parameters (inverse of Hessian matrix):

$$Var(\hat{\theta}) = \left(\mathbb{E}\left[\frac{\partial l}{\partial \theta}\frac{\partial l}{\partial \theta'}\right]\right)^{-1}.$$

2.1.2. Realized volatility

- Realized volatility uses high-frequency data to measure variance.
- Simple example: Suppose we sample returns in n intervals of length h, so that $T = nh$.
- Assume that log prices, p_t (with dividends reinvested), are a geometric Brownian motion with **constant volatility**

$$\mathrm{d}p_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t,$$

where W_t is a standard Brownian motion.

- We have for the h-period return:

$$r_{t+h,h} = p_{t+h} - p_t \sim N\left(\mu h, \sigma^2 h\right),$$

for any $h > 0$.

- The standard maximum likelihood estimators of μ and σ^2 are

$$\begin{aligned}\hat{\mu} &= \frac{1}{nh}\sum_{k=1}^{n} r_{kh,h}, \\ \hat{\sigma}^2 &= \frac{1}{nh}\sum_{k=1}^{n}\left(r_{kh,h} - \hat{\mu}h\right)^2.\end{aligned}$$

- If we compute the asymptotic distribution of the estimators:

$$\begin{aligned}\sqrt{T}\,(\hat{\mu}-\mu) &\to^d N(0,\sigma^2),\\ \sqrt{n}\left(\hat{\sigma}^2-\sigma^2\right) &\to^d N(0,2\sigma^4).\end{aligned}$$

- Key insight: The asymptotic variance of $\hat{\sigma}^2$ only depends on n and not on T.

- If we fix T and sample the data more frequently, then $n \to \infty$, and we can get an arbitrarily precise estimator of σ^2, no matter how small T.

- In contrast, to get a more precise estimator of μ we need to increase T.

- Logic explained further in Merton (1980).

- Now suppose **volatility is time-varying** and log prices are a semi-martingale

$$\mathrm{d}p_t = \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

where σ_t is a predictable process and square integrable,

$$E\left(\int_0^t \sigma_s^2 \mathrm{d}s\right) < \infty.$$

- Then, with $r_{t+1} = p_{t+1} - p_t$, and conditioning on the path of σ_t we have

$$r_{t+1} \mid \{\sigma_{t+\tau}\}_{\tau\in[0,1]} \sim N\left(\mu, \int_0^1 \sigma_{t+\tau}^2 \mathrm{d}\tau\right),$$

where $\int_0^1 \sigma_{t+\tau}^2 \mathrm{d}\tau$ is referred to as the integrated variance.

- In the absence of jumps, the **quadratic variation equals the integrated variance**. In this case, it holds:

$$\sum_{k=1}^{1/h} r_{t+kh,h}^2 \to QV(t,t+1) = IV(t,t+1) = \int_0^1 \sigma_{t+\tau}^2 \mathrm{d}\tau,$$

as $h \to 0$. Applied in finance by Andersen and Bollerslev (1998).

- Hence, with frequent sampling, we can –asymptotically– measure integrated variance without error from realized variance by summing squared returns.

- In practice, this means that we estimate realized volatility, say, every day, using data sampled intra-daily. We then treat the daily realized volatilities as an observed random variable for which we can use standard ARMA machinery for forecasting.

- Note: We did not need to subtract the mean before squaring returns.

- Why? Squared mean return goes to zero faster than variance when $h \to 0$, so we can ignore the squared mean term asymptotically. To see this, consider the case of i.i.d. returns

$$\begin{aligned} E\left(\sum_{k=1}^{1/h} r_{t+kh,h}^2\right) &= h^{-1}E\left(r_{t+h,h}^2\right) \\ &= h^{-1}Var\left(r_{t+h,h}\right) + h^{-1}\left(\mu h\right)^2 \\ &= Var\left(r_{t+1,1}\right) + h\mu^2. \end{aligned}$$

- In the presence of jumps, the realized variance estimator converges to the **quadratic variation**, which now **equals the sum of the integrated variance and the sum of squared jumps** from $[t, t+1]$.

- More formally
$$\mathrm{d}p_t = \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t + \xi_t \mathrm{d}q_t,$$
where q_t is a Poisson process with jump intensity λ_t.
ξ_t measures the impact of the jump on prices (jump size).

- The quadratic variation then equals
$$QV(t, t+1) = \int_0^1 \sigma^2_{t+\tau} \mathrm{d}\tau + \sum_{\tau \in [0,1]} J^2_{t+\tau} = IV(t, t+1) + \sum_{\tau \in [0,1]} J^2_{t+\tau}.$$

- You can decompose the quadratic variation into integrated variance and the jump component using the bipower variation as introduced in Bandorff-Nielson and Shephard (2004).

- See Andersen and Benzoni (2009) for a review of the techniques and literature.

- Although the theory calls for sampling at very high frequencies, intra-day returns are affected by microstructure noise:

    1. Bid/ask "bounce" - transaction prices may jump back and forth between bid and ask, depending on the sequence of buy and sell orders arriving in the market, possibly without having any effect on the mid-point price. This "volatility" is not the volatility we want to measure. Bid-ask bounce introduces negative autocorrelation.
    2. Price discreteness (depends on tick size).

- Microstructure noise is effectively measurement error, which leads to a bias in realized volatility.
- The bias gets worse the higher the sampling frequency.
- Illustration from Hansen and Lunde (2006)

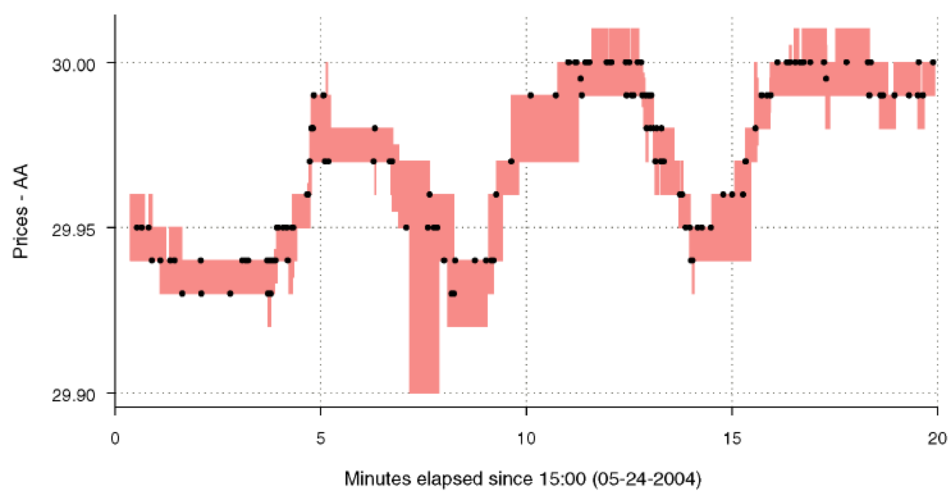


Figure 2. Bid and Ask Quotes (defined by the shaded area) and Actual Transaction Prices (●) Over Three 20-Minute Subperiods on April 24, 2004 for AA.

- One solution is to sample every 5 minutes, for instance, to avoid the discreteness.

2.1.3. Implied volatility

- In a standard Black-Scholes model, the price of a European option on a stock depends on
  - Moneyness (strike price relative to current stock price).
  - Maturity of the option.
  - Interest rate.
  - Volatility of the stock.
- Given that we observe option prices, moneyness, interest rates, and the maturity of options, we can compute the implied volatility.
- In a Black-Scholes model, each level of moneyness should give you the **same** implied volatility number, because it always is the volatility on the same stock.
- In practice, this is not the case, see for instance Dumas, Fleming, and Whaley (1998)

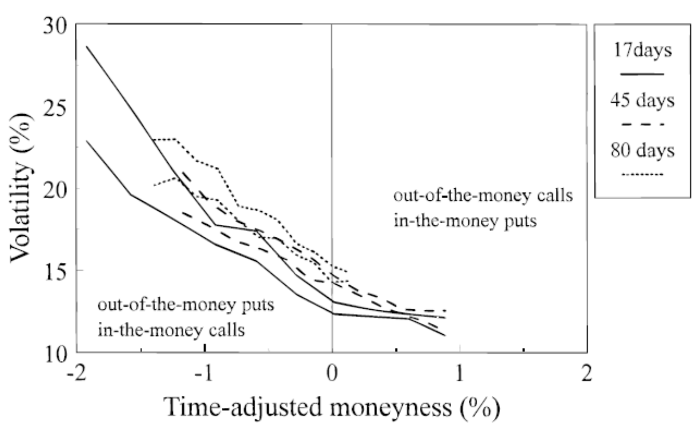


Figure 1. Black–Scholes implied volatilities on April 1, 1992. Implied volatilities are computed from S&P 500 index call option prices for the April, May and June 1992 option expirations. The lower line of each pair is based on the option's bid price, and the upper line is based on the ask. Time-adjusted moneyness is defined as $[X/(S - PVD) - 1]/\sqrt{T}$, where S is the index level, PVD is the present value of the dividends paid during the option's life, X is the option's exercise price, and T is its number of days to expiration.

- This hockey-stick line is called the volatility "smile" or "smirk"

- Implied vol is higher for out-of-the-money puts (they are expensive) and for longer-dated options.
- Implied volatility is negatively related to the level of the index

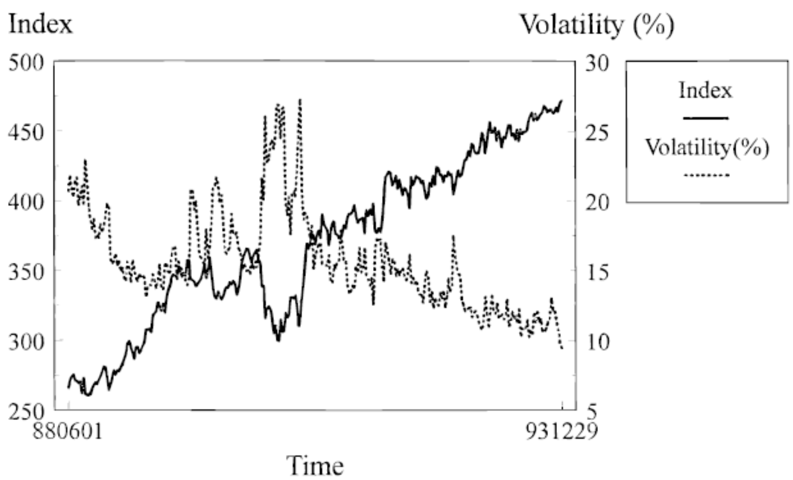
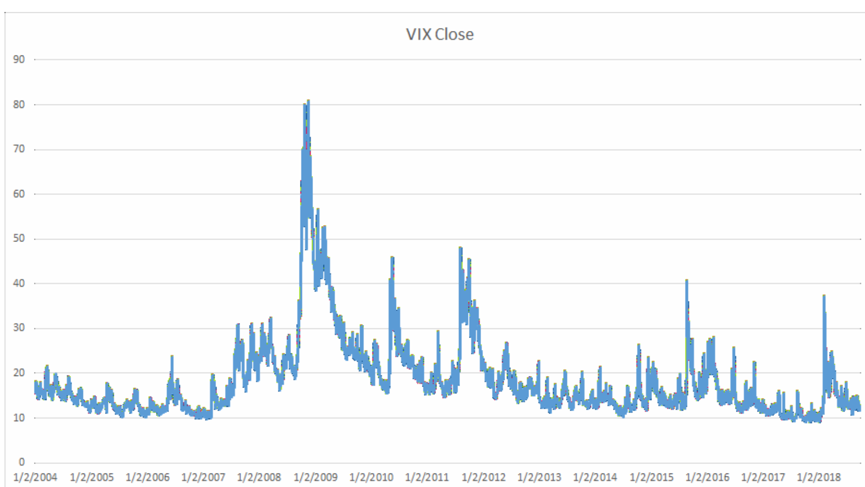


Figure 2. S&P 500 index level and Black–Scholes implied volatility each Wednesday during the period of June 1988 through December 1993.

- The VIX is a volatility index, computed using option prices on the S&P500 index across various strikes.
    - See CBOE for details on VIX construction.



- Investors can trade VIX options and VIX futures on the CBOE to take bets on aggregate volatility. In addition, there are ETFs that track VIX.

- Note: As the VIX is computed from option prices, it is the risk-neutral expectation of volatility (under the Q measure, not under the true P measure).

2.1.4. Text-based volatility measures: NVIX

- Since we do not have options data far back, we cannot compute the VIX prior to 1986. And although we can, in principle, compute realized variance for a long sample, the microstructure issues are more severe going back in time.

- Manela and Moreira (2017) construct a volatility measure using text data that starts in 1890. They call it news-implied volatility index or NVIX.

- An additional advantage of using text data is that we may get a sense what drives the VIX.

- Procedure:

  - Use all front-page articles from the WSJ from July 1889 until December 2009.

  - Break the sample into three parts

    1. 1996-2009 (training sub-sample): Link news data to implied volatility (VIX).
    2. 1986-1995 (test sub-sample): Used for out-of-sample tests and model fit.

3. 1889-1986 (predict sub-sample): This is the prediction sample for which the VIX is not available. Note that this out-of-sample period is very long.

- A month of text is summarized by a vector, x_t, of "n-gram frequencies"

$$x_{t,i} = \frac{\text{Appearances of n-gram } i \text{ in month } t}{\text{Total n-grams in month } t}$$

- Predict the VIX, v_t, using a linear regression

$$v_t = w_0 + w' x_t + e_t.$$

- We cannot estimate w precisely using OLS as x_t is very high dimensional. There are only 168 observations in the training sample.

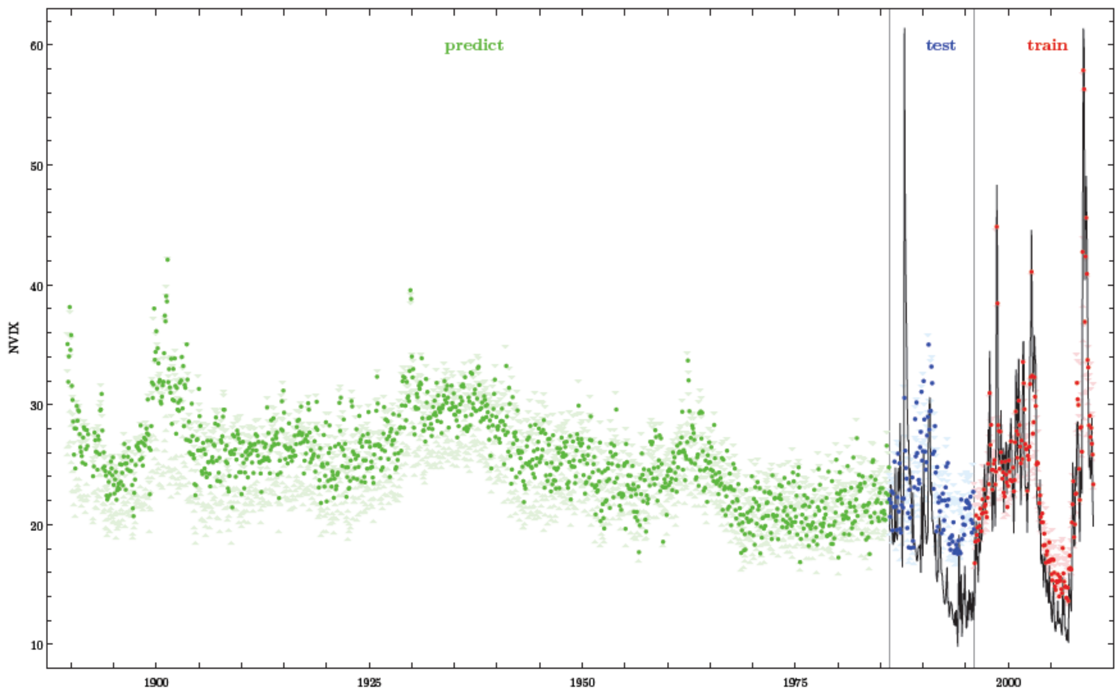
- Manela and Moreira (2017) propose to use "Support Vector Regressions." In this case, we choose w to minimize

$$\sum_{t \in train} g_\epsilon(v_t - w_0 - w' x_t) + c w' w,$$

  where $g_\epsilon(e) = \max\{0, |e| - \epsilon\}$, that is, a function that ignores errors smaller than ϵ and c is a regularization parameter (much like in ridge regressions). Selects a small number of n-grams (support vectors) and ignores the rest.

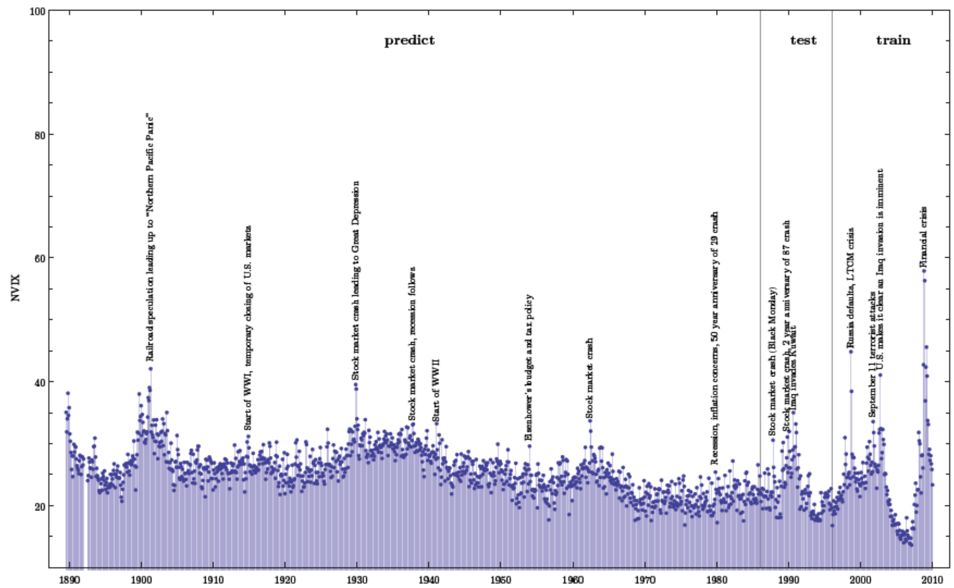
- The resulting NVIX series

Figure 1: News-Implied Volatility 1890–2009



Solid line is end-of-month CBOE volatility implied by options VIX_t. Dots are news implied volatility (NVIX) $VIX_t = w_0 + \mathbf{w} \cdot \mathbf{x}_t$, where $x_{t,i}$ are appearances of n-gram i in month t scaled by total month t n-grams, and $\mathbf{w}$ is estimated with a support vector regression. The *train* subsample, 1996 to 2009, is used to estimate the dependency between news data and implied volatility. The *test* subsample, 1986 to 1995, is used for out-of-sample tests of model fit. The *predict* subsample includes all earlier observations for which options data, and hence VIX is not available. Light-colored triangles indicate a nonparametric bootstrap 95% confidence interval around $\widehat{VIX}$ using 1000 randomizations. These show the sensitivity of the predicted values to randomizations of the *train* subsample.

Figure 2: News-Implied Volatility Peaks by Decade



We describe NVIX peak months each decade by reading the front page articles of *The Wall Street Journal* and cross-referencing with secondary sources when needed. Many of the market crashes are described in Mishkin and White (2002). See also Noyes (1909) and Shiller and Feltus (1989).

- Using realized volatility instead of VIX to test whether the predictive power holds up over the long sample, the results are quite stable:

Table 2

Out-of-sample realized volatility prediction using news.

Reported are model fit statistics repeating the estimation procedure over the same train subsample as before, only replacing implied volatility (VIX) with realized volatility as the dependent variable of the Support Vector Regression (SVR) Eq. (2). The train subsample, 1996–2009, is used to estimate the dependency between monthly news data and realized volatility. The test subsample, 1986 to 1995, is used for out-of-sample tests of model fit. The predict subsample includes all earlier observations for which options data and, hence, VIX are not available. *RMSE* SVR is root mean square error of the SVR. R^2 SVR is one less the prediction error's variance as a fraction of actual realized volatility's variance. *RMSE* Reg and R^2 Reg are estimated from a subsequent univariate Ordinary Least Squares (OLS) regression of actual realized volatility on realized volatility implied by news.

Subsample	*RMSE* SVR	R^2 SVR	*RMSE* Reg	R^2 Reg	Correlation	Observations
Train	3.38	90.69	2.62	92.70	96.28	168
Test	9.61	20.24	9.08	20.35	45.11	119
Predict	10.68	13.58	8.50	15.99	39.98	1150

2.2. Variance risk premium

2.2.1. Measuring the VRP at different horizons

- Given the counter-cyclical dynamics of volatility, it is natural to conjecture that risk-averse investors require compensation for being exposed to volatility shocks, and to ask how highly variance risk is priced.

- A direct way to trade variance risk is using variance swaps.

- The payoff of a variance swap at time t with maturity m is

$$\text{Variance Swap Payoff}_t^m = \sum_{j=t+1}^{t+m} r_j^2 - VS_t^m,$$

where r_j is the log return on the index at date j and VS_t^m the price of the variance swap. A period corresponds to a day.

- Fairly valued variance swaps have prices VS_t^m at origination:

$$VS_t^m = E_t^Q \left[\sum_{j=t+1}^{t+m} r_j^2 \right],$$

equal to the Q-expectations of future variance (quadratic variation).

- **Variance risk premium** is the difference between the P and Q expectations of future variance.

- The P-expectation of future variance can be constructed using a GARCH model or high-frequency data. This is a very direct way to measure the variance risk premium. The only downside is that the sample where variance swap prices are available is fairly short and that the market may be illiquid initially.

- Dew-Becker, Giglio, Le, and Rodriguez (2017) study the returns of variance swaps.
- They also define the forward price of variance as

$$F_t^m = VS_t^m - VS_t^{m-1}.$$

The m-month variance swap is the market's risk-neutral expectation of realized variance m months in the future at the end of month t. $F_t^0 = RV_t$. $F_t^1 = VS_t^1$.

- The forward prices of variance swaps

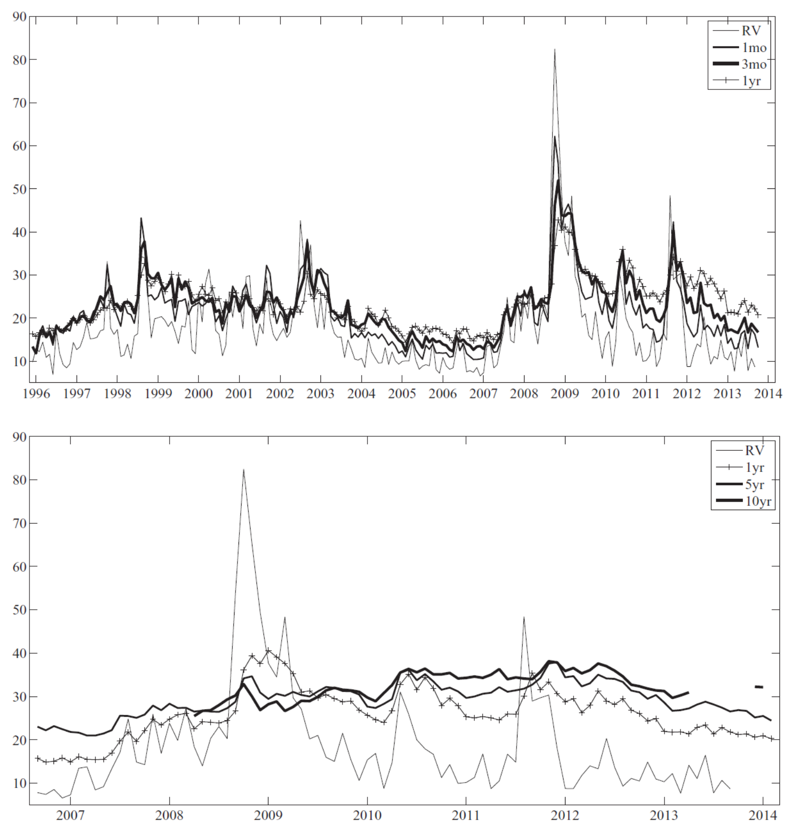


Fig. 1. Time series of forward variance claim prices. The figure shows the time series of forward variance claim prices of different maturities. For readability, each line plots the prices in annualized volatility terms, $100 \times \sqrt{12 \times F_t^n}$, for a different n. The top panel plots forward variance claim prices for maturities of one month, three months, and one year. The bottom panel plots forward variance claim prices for maturities of 1 year, 5 years and 10 years. Both panels also plot annualized realized volatility, $100 \times \sqrt{12 \times F_t^0}$.

- Term structure of variance forward prices usually weakly upward sloping. The curve inverts in times of distress, with RV spiking. In those periods, the market expects the variance to normalize (go down) in future.

- Return on a variance forward related to slope of forward curve:

$$R_{t+1}^{m} = \frac{F_{t+1}^{m-1} - F_{t}^{m}}{F_{t}^{m}} = \frac{E_{t+1}^{Q}[RV_{t+m}] - E_{t}^{Q}[RV_{t+m}]}{E_{t}^{Q}[RV_{t+m}]}$$

  Measures change in expectations about volatility at its maturity.

- Sharpe ratios of variance swap returns across maturities.

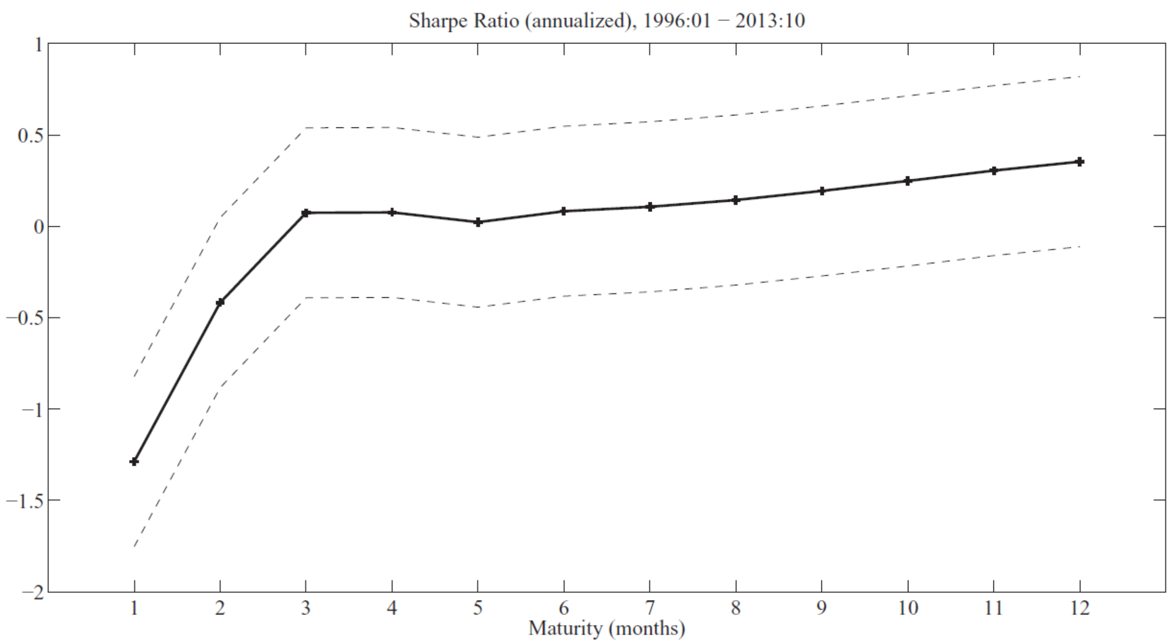


Fig. 5. Annualized Sharpe ratios for forward variance claims. The figure shows the annualized Sharpe ratio for the forward variance claims. The returns are calculated assuming that the investment in an n-month variance claim is rolled over each month. Dotted lines represent 95% confidence intervals. All tests for the difference in Sharpe ratio between the one-month variance swap and any other maturity confirm that they are statistically different with a p-value of 0.03 (for the second month) and < 0.01 (for all other maturities). The sample used is 1996–2013.

- Negative average returns on variance forwards = investors are willing to pay a large risk premium to hedge realized volatility (SR$<$-1.0). Equivalently, investors who are providing this insurance earn a large risk premium (SR $>$ 1.0).

- This suggests that the variance risk premium is large for short-run variance risk.

- However, the variance risk premium declines quickly with maturity.

- Investors **do not require compensation** for exposure to news about future expected volatility at horizons beyond 2 months. They are not willing to pay a risk premium to hedge this risk.

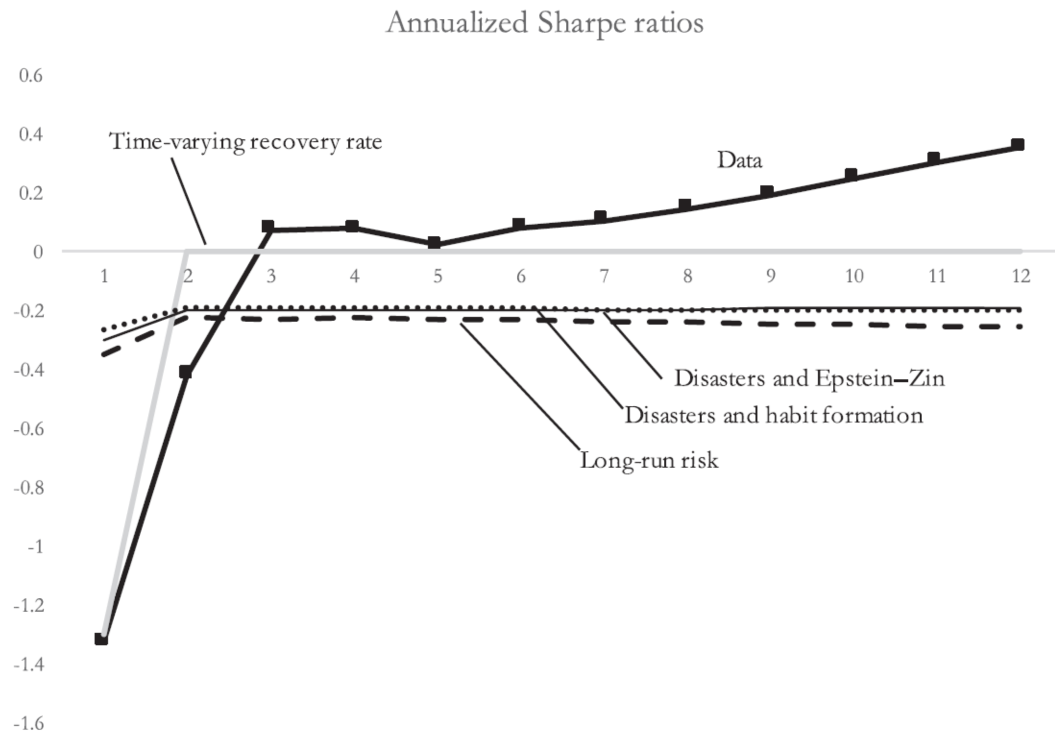
- Put differently, insuring volatility risk at horizons beyond 2 months is **free**!

- Only the transitory part of realized variance appears to be priced in the 1996-2014 data.

- This fact poses a big challenge to structural asset pricing models (see next slide).

- The authors also suggest that their findings imply that aggregate volatility shocks are unlikely to be a major driver of business cycles or consumer welfare. More on this in section 4 of these notes.

- SR on variance forwards in theoretical AP models



- Two observations:

  1. The models cannot replicate the very high Sharpe ratio for short-maturity variance claims, and the downward sloping Sharpe ratios as maturity increases.
  2. The models all rely on persistent state variables, which leads to non-trivial variance risk premia at longer maturities, which is not supported empirically.

2.2.2. Predicting Stock Returns with the VRP

- Bollerslev, Tauchen, and Zhou (2009) study the predictability of the variance risk premium for future stock market returns.

- It is natural to conjecture that times in which the variance risk premium is high are times in which the equity risk premium is also high.

  - In a long-run risk model, for example, shocks to volatility are priced and increase the equity risk premium (when IES$>$1 and risk aversion $>$1).

  - When the volatility of volatility is time-varying, the variance risk premium is time-varying and strongly correlated with the equity risk premium.

- The variance risk premium is defined as

$$VRP_t = IV_t - RV_t,$$

  where IV_t is the implied variance (=squared VIX) and RV_t the realized variance.

- The implicit assumption here is that

$$E_t^P(RV_{t+1}) \simeq RV_t.$$

- To construct RV_t, they sum 5-minute squared returns during a month. This gives $22 \times 78 = 1,716$ squared returns for a typical month.

- The sample period is from January 1990 to December 2007.

- The dynamics of implied variance, realized variance, and the variance risk premium:

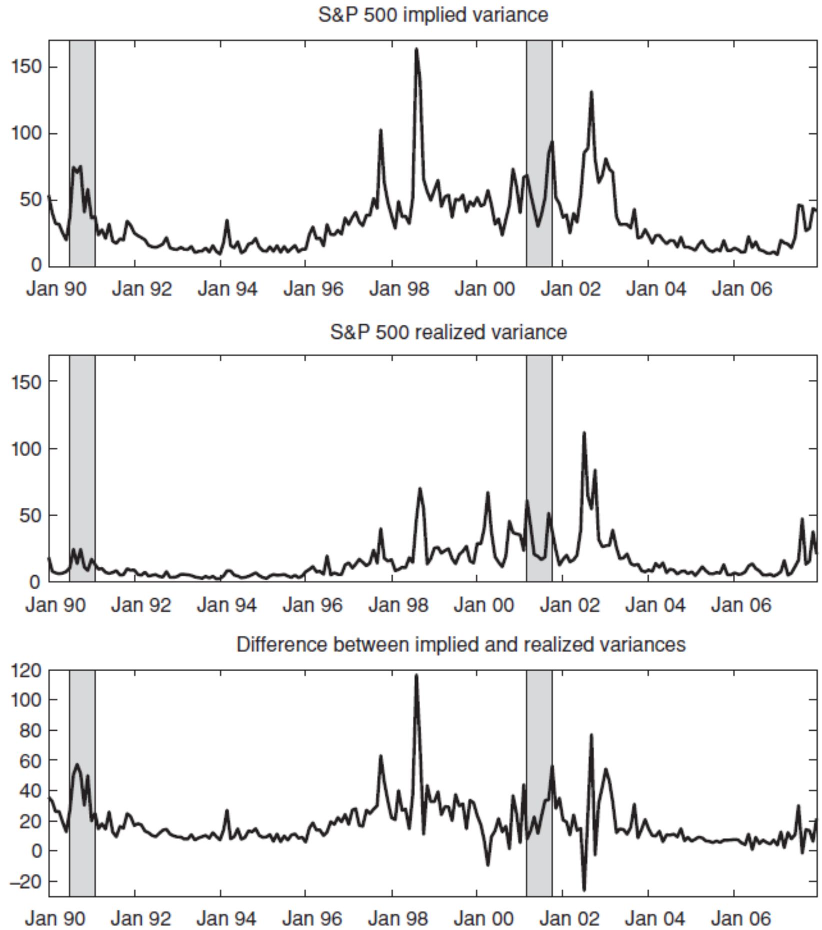


Figure 2
Implied and realized variances and variance risk premium
This figure plots the implied variance (the top panel), the realized variance (the middle panel), and the difference (the bottom panel) for the S&P 500 market index from January 1990 to December 2007. The shaded areas represent NBER recessions.

- Strong predictability for monthly and quarterly market returns, weaker for annual returns

Table 4
Quarterly return regressions

	Simple									Multiple				
Constant	−2.08	0.24	6.60	92.41	73.35	20.63	7.39	6.92	5.53	101.89	−4.12	85.93	100.06	98.21
	(−0.56)	(0.06)	(1.60)	(2.17)	(1.81)	(1.32)	(1.24)	(2.18)	(1.54)	(2.40)	(−1.00)	(1.67)	(1.93)	(2.18)
$IV_t - RV_t$	0.47									0.58	0.51		0.59	0.70
	(2.86)									(3.43)	(3.02)		(3.38)	(4.01)
IV_t		0.19												
		(1.41)												
RV_t			0.00											
			(0.00)											
$\log(P_t/E_t)$				−2.28						−2.82		−2.11	−2.77	−2.95
				(−1.97)						(−2.42)		(−1.54)	(−1.98)	(−2.33)
$\log(P_t/D_t)$					−1.42									
					(−1.62)									
$DFSP_t$						−1.39								
						(−0.90)								
$TMSP_t$							−0.46							4.08
							(−0.17)							(1.42)
$RREL_t$								3.27						6.39
								(0.88)						(1.56)
CAY_t									3.23		3.52	1.08	0.74	
									(1.78)		(1.99)	(0.53)	(0.37)	
Adj. R^2 (%)	6.82	2.49	−0.47	6.55	4.19	1.18	−0.43	0.43	4.13	16.76	11.87	7.21	17.42	19.74

The sample period extends from January 1990 to December 2007. All of the regressions are based on overlapping monthly observations. Robust t-statistics accounting for the overlap following Hodrick (1992) are reported in parentheses. All variable definitions are identical to Tables 1 and 3.

Table 5
Annual return regressions

	Simple									Multiple				
Constant	4.62	7.62	9.49	78.47	79.83	15.59	5.37	7.29	5.42	81.00	1.91	52.85	55.11	74.04
	(1.50)	(2.44)	(3.20)	(2.05)	(2.17)	(1.13)	(0.90)	(2.33)	(1.47)	(2.15)	(0.53)	(1.03)	(1.08)	(1.88)
$IV_t - RV_t$	0.12									0.19	0.18		0.20	0.33
	(1.00)									(1.68)	(1.51)		(1.74)	(2.96)
IV_t		−0.02												
		(−0.21)												
RV_t			−0.17											
			(−1.20)											
$\log(P_t/E_t)$				−1.90						−2.06		−1.24	−1.40	−2.14
				(−1.80)						(−2.00)		(−0.91)	(−1.03)	(−1.92)
$\log(P_t/D_t)$					−1.55									
					(−1.92)									
$DFSP_t$						−0.87								
						(−0.64)								
$TMSP_t$							0.88							4.53
							(0.35)							(1.69)
$RREL_t$								4.09						6.29
								(1.11)						(1.75)
CAY_t									3.48		3.62	2.13	2.12	
									(1.99)		(2.12)	(0.99)	(0.99)	
Adj. R^2 (%)	1.23	−0.37	2.89	16.34	19.53	1.79	0.01	4.54	18.15	20.12	21.18	21.46	25.52	32.58

The sample period extends from January 1990 to December 2007. All of the regressions are based on overlapping monthly observations. Robust t-statistics accounting for the overlap following Hodrick (1992) are reported in parentheses. All variable definitions are identical to Tables 1 and 3.

- Further evidence:
  - Bekaert and Hoerova (2014) use various models for the expectation of realized variance. Continue to find predictability for returns by the alternative VRP measures.
  - International equities: Bollerslev, Marrone, Xu, and Zhou (2014) ⇒ Global variance risk premium.
  - Currency returns: Londono and Zhou (2017).

2.3. Volatility and the cross-section of expected returns

- Instead of using variance swaps, we can use the cross-section of stock returns to estimate the price of variance risk, see Ang, Hodrick, Xing, and Zhang (2006).

- Advantage: Long sample.

- Disadvantages:

  1. We do not get the rich term structure implications of variance swaps.

  2. There are many other factors affecting the cross-section of expected returns besides variance risk.

- Two sorts:

  1. Exposure to aggregate volatility shocks.

  2. Idiosyncratic risk level.

- Regression for aggregate volatility

$$r_t^i = \beta_0 + \beta_{MKT}^i MKT_t + \beta_{\Delta VIX}^i \Delta VIX_t + \epsilon_t^i,$$

and sort stocks on $\beta_{\Delta VIX}^i$.

Table I

Portfolios Sorted by Exposure to Aggregate Volatility Shocks

We form value-weighted quintile portfolios every month by regressing excess individual stock returns on ΔVIX, controlling for the MKT factor as in equation (3), 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 and B/M reports the average book-to-market ratio. 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–French (1993) three-factor model. The pre-formation betas refer to the value-weighted $\beta_{\Delta VIX}$ or β_{FVIX} within each quintile portfolio at the start of the month. We report the pre-formation $\beta_{\Delta VIX}$ and 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} factor loadings 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 $FVIX$ factor that mimics aggregate volatility risk, following the regression in equation (6). The row labeled "Joint test p-value" reports a Gibbons, Ross and Shanken (1989) test for the alphas equal to zero, and a robust joint test that the factor loadings are equal to zero. Robust Newey–West (1987) t-statistics are reported in square brackets. The sample period is from January 1986 to December 2000.

								Factor Loadings			
Rank	Mean	Std. Dev.	% Mkt Share	Size	B/M	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}
1	1.64	5.53	9.4%	3.70	0.89	0.27	0.30	−2.09	−2.00	−0.033	−5.06
						[1.66]	[1.77]				[−4.06]
2	1.39	4.43	28.7%	4.77	0.73	0.18	0.09	−0.46	−0.42	−0.014	−2.72
						[1.82]	[1.18]				[−2.64]
3	1.36	4.40	30.4%	4.77	0.76	0.13	0.08	0.03	0.08	0.005	−1.55
						[1.32]	[1.00]				[−2.86]
4	1.21	4.79	24.0%	4.76	0.73	−0.08	−0.06	0.54	0.62	0.015	3.62
						[−0.87]	[−0.65]				[4.53]
5	0.60	6.55	7.4%	3.73	0.89	−0.88	−0.53	2.18	2.31	0.018	8.07
						[−3.42]	[−2.88]				[5.32]
5-1	−1.04					−1.15	−0.83				
	[−3.90]					[−3.54]	[−2.93]				
Joint test p-value						0.01	0.03				0.00

- Stocks that hedge against aggregate volatility shocks earn lower average returns and alphas (83bps per month 3-FF alpha). This makes considerable sense.

- Papers studying how aggregate volatility is priced in the cross-section:
  - Bansal, Kiku, Shaliastovich, and Yaron (2014).
  - Campbell, Giglio, Polk, and Turley (2018).

- For idiosyncratic risk, take the residual of the Fama and French 3-factor model:

$$r_t^i = \alpha^i + \beta_{MKT}^i MKT_t + \beta_{SMB}^i SMB_t + \beta_t^i HML_t + \epsilon_t^i,$$

estimated using daily data. Idiosyncratic risk is measured as $\sqrt{var(\epsilon_t^i)}$.

Panel B: Portfolios Sorted by Idiosyncratic Volatility Relative to FF-3							
1	1.04	3.83	53.5%	4.86	0.85	0.11	0.04
						[1.57]	[0.99]
2	1.16	4.74	27.4%	4.72	0.80	0.11	0.09
						[1.98]	[1.51]
3	1.20	5.85	11.9%	4.07	0.82	0.04	0.08
						[0.37]	[1.04]
4	0.87	7.13	5.2%	3.42	0.87	−0.38	−0.32
						[−2.32]	[−3.15]
5	−0.02	8.16	1.9%	2.52	1.10	−1.27	−1.27
						[−5.09]	[−7.68]
5-1	−1.06					−1.38	−1.31
	[−3.10]					[−4.56]	[−7.00]

- Portfolio of stocks with high idiosyncratic risk has lower average returns, after controlling for exposure to the FF3 factors.

- Effects are large: 1.3% per month 3-factor alpha.

- Effects cannot be explained by exposure to aggregate volatility risk (double-sorts on aggr. volatility beta and idio. vol. level).

- In standard models, there is no compensation for idiosyncratic risk.

- In models with frictions, e.g., information frictions, where investors cannot fully diversify away idiosyncratic risk, one would expect a *positive* relationship between idiosyncratic risk and expected return. Data show the **opposite**.

- Ang, Hodrick, Xing, and Zhang (2009) study potential explanations for the U.S. and provide additional international evidence.
- Fama-MacBeth regression of excess returns on lagged idiosyncratic volatility, factor betas, and firm characteristics:

Table 2
Idiosyncratic volatility and expected returns in G7 countries

	Canada	France	Germany	Italy	Japan	U.K.	U.S.
Panel A: USD-denominated returns							
Constant	1.723	0.602	0.753	0.425	0.948	0.480	1.746
	[3.68]	[1.13]	[1.87]	[0.76]	[1.25]	[1.03]	[3.83]
W-FF idiosyncratic volatility	−1.224	−1.439	−2.003	−1.572	−1.955	−0.871	−2.014
	[−2.46]	[−2.14]	[−3.85]	[−2.10]	[−5.18]	[−2.54]	[−6.67]
$\beta(MKT^W)$	0.344	0.059	0.277	−0.083	0.323	0.178	0.376
	[2.20]	[0.44]	[1.93]	[−0.32]	[3.12]	[1.46]	[4.52]
$\beta(SMB^W)$	0.009	0.015	−0.083	0.116	0.050	0.032	−0.049
	[0.12]	[0.17]	[−0.82]	[0.56]	[0.76]	[0.42]	[−1.19]
$\beta(HML^W)$	−0.070	−0.069	0.076	−0.221	−0.025	−0.077	−0.051
	[−0.95]	[−0.94]	[1.00]	[−1.98]	[−0.35]	[−1.30]	[−1.69]
Size	−0.253	−0.067	−0.044	−0.031	−0.132	−0.058	−0.157
	[−4.81]	[−1.08]	[−1.09]	[−0.47]	[−1.72]	[−1.16]	[−3.14]
Book-to-market	0.369	0.569	0.176	0.239	0.550	0.365	0.282
	[3.68]	[4.59]	[1.35]	[1.48]	[3.84]	[4.46]	[3.87]
Lagged return	0.014	0.001	0.003	0.001	−0.011	0.012	−0.001
	[3.57]	[0.10]	[1.01]	[0.15]	[−2.85]	[4.07]	[0.28]
Adjusted R^2	0.118	0.108	0.114	0.147	0.124	0.078	0.046
Percentiles of W-FF idiosyncratic volatility							
25th Percentile	20.8	21.4	16.3	21.5	23.1	13.9	25.0
75th Percentile	46.0	39.2	34.8	38.4	39.6	31.3	61.1
Economic effect of moving from the 25th to the 75th W-FF idiosyncratic volatility percentiles							
25% → 75%	−0.31%	−0.26%	−0.37%	−0.27%	−0.32%	−0.15%	−0.73%

Panel B: Local-currency-denominated returns						
Constant	1.730	0.319	0.554	0.653	0.657	0.513
	[3.70]	[0.56]	[1.34]	[1.10]	[0.94]	[1.11]
L-FF idiosyncratic volatility	−1.332	−1.057	−1.769	−1.865	−2.035	−0.934
	[−2.59]	[−1.64]	[−3.38]	[−2.76]	[−5.89]	[−2.63]
$\beta(MKT^W)$	0.422	0.133	0.413	0.014	0.999	0.525
	[2.64]	[0.71]	[2.13]	[0.05]	[5.76]	[3.59]
$\beta(SMB^W)$	0.123	−0.044	0.037	−0.011	−0.016	−0.048
	[1.30]	[−0.45]	[0.37]	[−0.07]	[−0.15]	[−0.54]
$\beta(HML^W)$	−0.077	0.114	0.178	−0.126	0.012	−0.022
	[−0.82]	[1.31]	[2.10]	[−1.11]	[0.10]	[−0.38]
Size	−0.254	−0.041	−0.039	−0.080	−0.143	−0.090
	[−4.84]	[−0.65]	[−0.95]	[−1.19]	[−2.01]	[−1.72]
Book-to-market	0.406	0.571	0.147	0.253	0.552	0.321
	[3.68]	[4.74]	[1.03]	[1.77]	[3.94]	[4.04]
Lagged return	0.015	0.001	0.001	0.001	−0.011	0.012
	[3.69]	[0.29]	[0.42]	[0.16]	[−2.90]	[4.09]
Adjusted R^2	0.110	0.107	0.115	0.144	0.131	0.073

The table reports Fama-MacBeth (1973) regressions (Eq. (4)) for the individual G7 countries. We regress monthly excess firm returns on a constant; idiosyncratic volatility over the past month with respect to the W-FF model in Eq. (3); contemporaneous factor loadings, $\beta(MKT^W)$, $\beta(SMB^W)$, and $\beta(HML^W)$ with respect to the W-FF model; and firm characteristics at the beginning of the month. "Size" is the log market capitalization of the firm at the beginning of the month, "Book-to-market" is the book-to-market ratio available six months prior, and "Lagged return" is the firm return over the previous six months. We report the robust t-statistics in square brackets below each coefficient. The row "Adjusted R^2" reports the average of the cross-sectional adjusted R^2's. Each cross-sectional regression is run separately for each country using U.S. dollar-denominated firm excess returns in Panel A and local-currency-denominated firm excess returns in Panel B. In Panel A, we also report the 25th and 75th percentiles of each country's W-FF idiosyncratic volatility and compute the economic effect of moving from the 25th to the 75th percentile. For example, for Canada, a move from the 25th to the 75th percentile of W-FF idiosyncratic volatility would result in a decrease in a stock's expected return of $|-1.224| \times (0.460 - 0.208) = 0.31\%$ per month. The sample period is from January 1980 to December 2003 for all countries.

- Herskovic, Kelly, Lustig and Van Nieuwerburgh (2016) show that idiosyncratic volatilities of U.S. stocks are synchronized.
- A single factor, the common idiosyncratic volatility or CIV factor, explains 35% of the time-series variation in firm-level idiosyncratic return variance.
- Same is true for volatility of fundamentals, such as sales growth.
- Stocks with high CIV-beta have low average returns (panels A, B), even after controlling for MV-beta exposure (panels C, D) or idiosyncratic variance (appendix table A.3, not reported here):

Table 2
Portfolios formed on common idiosyncratic volatility (CIV)-beta.

The table reports average excess returns and alphas in annual percentages for portfolios sorted on the basis of monthly CIV-beta for the 1963–2010 sample. Panel A reports equally weighted average excess returns and alphas in one-way sorts using all Center for Research in Security Prices (CRSP) stocks. Panel B reports value-weighted averages in one-way sorts. Panel C shows equally weighted one-way sorts on CIV-beta that control for market variance (MV)-beta. Panel D shows equally weighted average excess returns in sequential two-way sorts on CIV-beta and MV-beta.

	CIV beta						
	1 (low)	2	3	4	5 (high)	5 − 1	$t(5-1)$
Panel A: One-way sorts on CIV-beta							
$\mathbb{E}[R]-r_f$	12.08	10.88	9.96	8.70	6.68	−5.41	−3.94
α_{CAPM}	5.38	5.07	4.55	3.24	0.61	−4.77	−3.52
α_{FF}	1.06	1.07	0.78	−0.07	−2.23	−3.29	−2.50
Panel B: One-way sorts on CIV-beta (value-weighted)							
$\mathbb{E}[R]-r_f$	9.41	7.04	5.77	5.82	3.87	−5.53	−3.15
α_{CAPM}	2.84	1.34	0.49	0.74	−1.72	−4.56	−2.65
α_{FF}	1.58	0.58	0.26	0.78	−1.59	−3.17	−1.84
Panel C: One-way sorts on CIV-beta controlling for MV-beta							
$\mathbb{E}[R]-r_f$	11.71	11.08	9.57	8.57	7.36	−4.35	−3.10
α_{CAPM}	4.87	5.04	3.99	3.21	1.74	−3.14	−2.38
α_{FF}	0.66	1.22	0.31	−0.19	−1.39	−2.05	−1.64
Panel D: Two-way sorts on CIV-beta and MV-beta							
1 (low)	10.04	10.36	8.48	7.66	6.16	−3.88	−2.52
2	12.28	10.08	9.24	9.54	8.36	−3.92	−2.25
3	12.51	11.09	9.88	8.50	6.80	−5.71	−3.38
4	12.88	11.43	9.98	7.76	7.46	−5.42	−3.29
5 (high)	10.86	12.46	10.26	9.37	8.04	−2.82	−1.39
5 − 1	0.81	2.10	1.78	1.71	1.88	–	–
$t(5-1)$	0.48	1.17	1.01	0.90	0.84	–	–

- CIV exposure captures a new source of systematic risk; paper shows it has a significantly negative price of risk $\lambda_{CIV} < 0$.

- CIV factor also helps price size-, BM-, EP-, momentum, and aggregate volatility beta-sorted portfolios.

- Traded version of the CIV factor has average return very similar to the estimated λ_{CIV}.

- Paper proposes model where firm-level dividend growth has an idiosyncratic shock component whose variance is common across firms, the CIV factor.

- Households face idiosyncratic consumption growth risk, due to incomplete markets. Shocks to household-specific consumption growth have time-varying volatility also driven by the CIV factor.

- Intuition: some idiosyncratic firm cash-flow risk is passed through from firms to the households (e.g., via compensation contracts or layoffs), and households require compensation to bear this risk.

- Periods of high CIV in firms' cash flows are periods where risk sharing is difficult, and the cross-sectional variance of individual consumption growth increases.

- Evidence from individual earnings data shows comovement between dispersion in labor income growth and CIV factor.

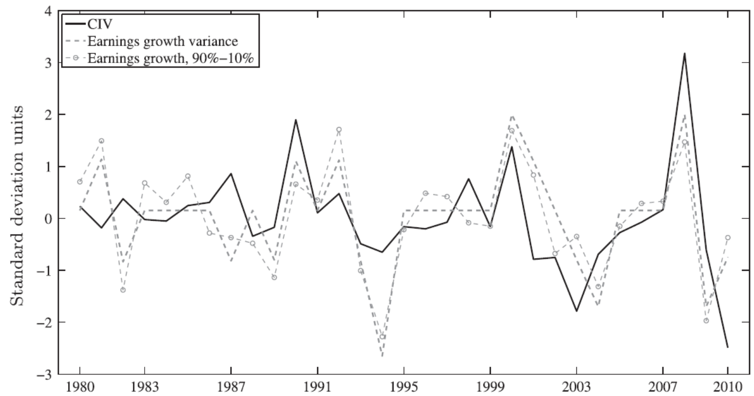


Fig. 6. Common idiosyncratic volatility (CIV) and dispersion in individual household income growth. The figure compares yearly changes in CIV with yearly changes in the standard deviation and interdecile range of the individual earnings growth distribution. CIV is the equal-weighted average of firm-level market model residual return variance each year. Individual earnings data are from the US Social Security Administration and summarized by Guvenen, Ozkan, and Song (2014). Each series is standardized to have equal mean and variance for ease of comparison.

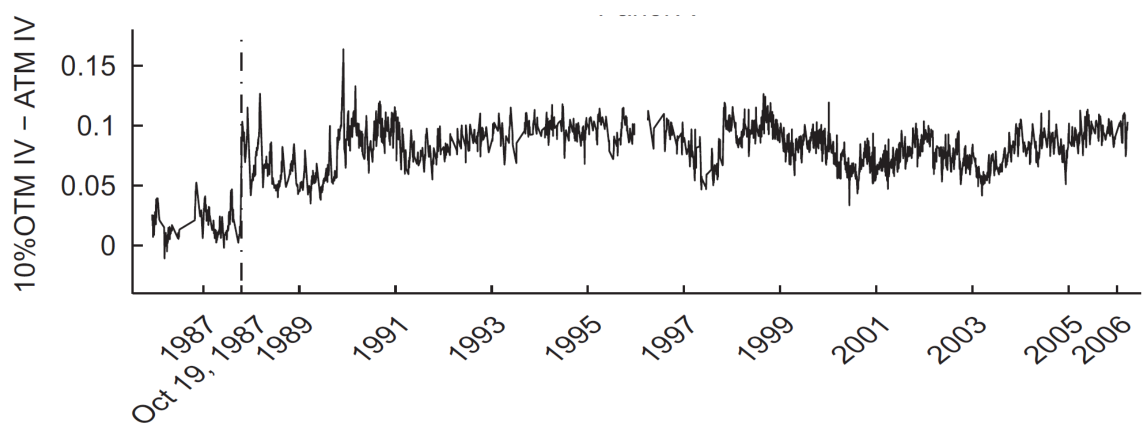
3. Interpreting the Facts

3.1. Consumption-based asset pricing models

- Growing literature tries to understand the dynamics of volatility and option prices.
- Traditional asset pricing models with interesting variance and variance risk premium dynamics:
  - Long-run risks: Drechsler and Yaron (2011). To match the dynamics of volatility and the variance risk premium, they add jumps in consumption, dividends, and the variance process.
  - Variable rare disasters: Gabaix (2012) and Seo and Wachter (2019).

3.2. Learning-based models

- Benzoni, Collin-Dufresne, and Goldstein (2011) provide a learning-based explanation of the change in the implied volatility curve following the 1987 stock market crash.
- Motivating figure:



- The implied volatility curve was reasonably flat before the crash.
- But the implied volatility curve steepened and remained steep for the next 20 years.

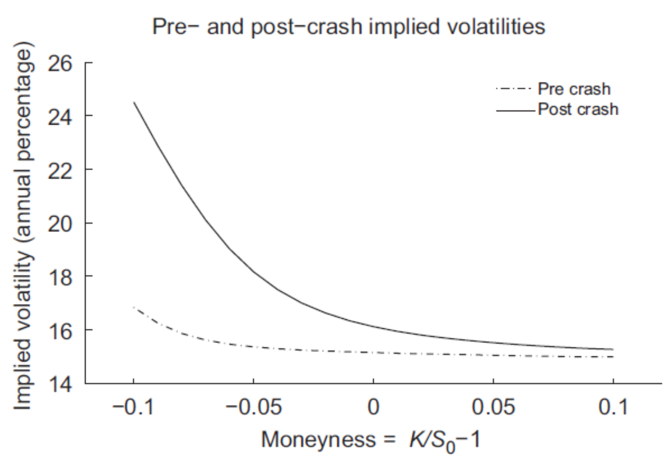


Fig. 2. The plots depict the model-implied volatility smirk pre- and post-1987 market crash for S&P 500 options with one month to maturity. The model coefficients are set equal to the baseline values given in Table 1.

- Arrival of a rare jump triggers the updating of agents' beliefs about the likelihood of future jumps, which produces a market crash, and a permanent shift in option prices.
- Similar intuition in Kozlowski, Veldkamp, and Venkateswaran (2020) to explain the secular stagnation after the 2007-09 Great Financial Crisis.
  - An extreme event was realized that was not yet in the econometrician's information set.
  - After observing this event, this permanently changed his beliefs about the distribution of macro outcomes.
  - Increase in tail risk. SKEW index uses out-of-the-money S&P500 options to back out the risk-neutral probability of a left-tail event. It increases in the GFC and **remains high**.

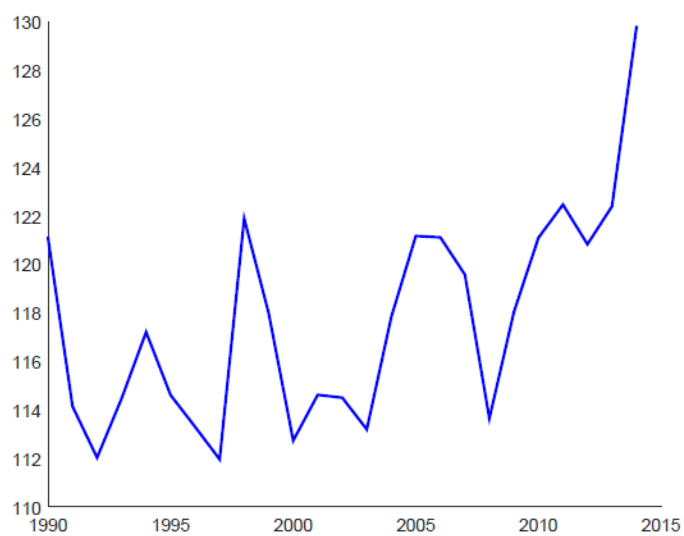


Figure 2: **The SKEW Index.**
A measure of the market price of tail risk on the S&P 500, constructed using option prices. Source: Chicago Board Options Exchange (CBOE). 1990:2014.

3.3. *Intermediary-based asset pricing models*

- Garleanu, Pedersen, and Poteshman (2009) develop an intermediary-based asset pricing model for option markets.
  - Imbalances in dealer inventory matter for option pricing
  - They relate "demand pressure" from end users to option puzzles such as the fact that OTM puts are expensive.
  - Model is related to the Vayanos and Vila (2021) model, which we will discuss next week in context of the term structure of interest rates.
- Similarly, Adrian and Shin (2010) link balance sheet variables of dealers to volatility and the variance risk premium.
- Coimbra and Rey (2019) develop a model with intermediaries that have heterogenous Value-at-Risk constraints.
- Outline of the Garleanu, Pedersen, and Poteshman (2009) model:
  - Derivatives are indexed by $i = 1, \ldots, I_t$ with prices $p_t = (p_t^i)_{i \in I_t}$.
  - Two groups investors:
    1. "End-users" of options: Demand $d_t = (d_t^i)_{i \in I_t}$.
    2. Dealers: Demand $q_t = (q_t^i)_{i \in I_t}$.
  - Options are in zero-net supply; market clearing implies:

$$d_t + q_t = 0$$

  - Dealers are competitive and infinitely-lived, risk averse with CARA preferences over consumption

$$\mathbb{E}_t \left[\sum_{v=t}^{\infty} \rho^{v-1} \frac{-\exp(-\gamma c_v)}{\gamma} \right]$$

- Modeling strategy: Use the dealer first-order condition to figure out option prices given market clearing and exogenous demand shocks. That is, compute prices that make the dealer position (observed in the data later on) optimal.
- Because dealers optimize, standard no-arbitrage relationships such as put-call parity hold.
- Note: If markets are complete, then demand pressure does not matter because dealers can hedge any risk in broader asset markets.
- Potential sources of market incompleteness:
  1. Discrete-time hedging (see Bertsimas, Kogan, and Lo (2000) for a precise analysis of the hedging error in discrete time).
  2. Jumps in the underlying.
  3. Stochastic-volatility risk.
- See Pan (2002) for a model with stochastic volatility and jumps, and an analysis of jump risk premia.
- Define the unhedgeable part of price changes as $\overline{p}^k_{t+1}$, where we take out the optimally-hedged stock return.
- Key result:

$$\frac{\partial p^i_t}{\partial d^j_t} = \gamma(R_f - 1)Cov_t(\overline{p}^i_{t+1}, \overline{p}^j_{t+1}),$$

  that is, the price effect for option i of demand pressure for option j depends on the covariance of the unhedgeable parts and the risk aversion of the dealer.

– Direct implication, own demand pressure is positive

$$\frac{\partial p_t^i}{\partial d_t^i} = \gamma(R_f - 1)Var_t(\overline{p}_{t+1}^i),$$

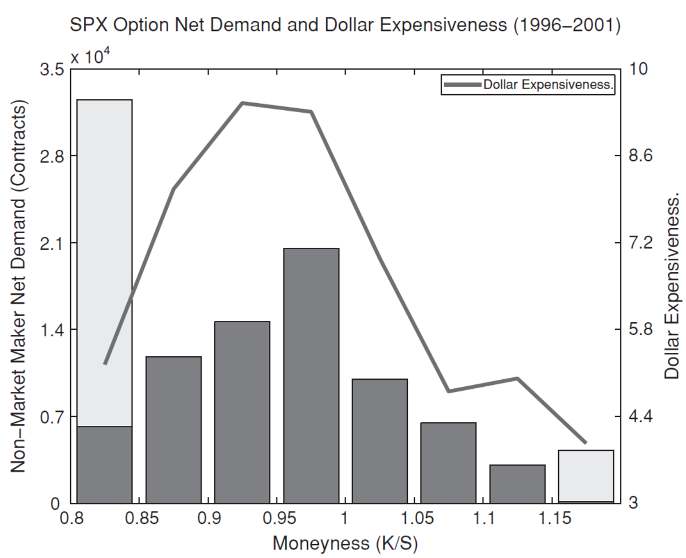
highlighting that price pressure is more pronounced for options with a large unhedgeable component. Also, when the dealer is more risk averse (e.g., when its capital position is weak), we expect to see more demand pressure.

- Fact #1: Net demand for out-of-the-money index puts is positive.

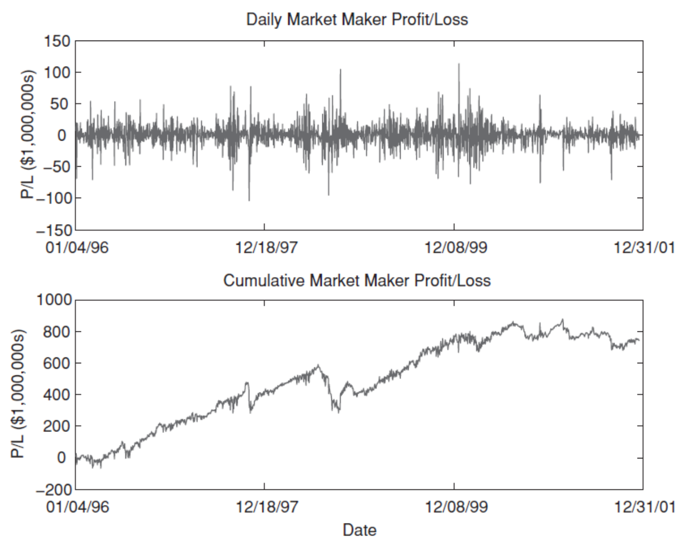
Net demand for options by end-users

Mat. range (cal. days)	Moneyness range (K/S)							
	0–0.85	0.85–0.90	0.90–0.95	0.95–1.00	1.00–1.05	1.05–1.10	1.10–1.15	1.15–2.00
	Panel A: SPX option non-market-maker net demand							
1–9	6,014	1,780	1,841	2,357	2,255	1,638	524	367
10–29	7,953	1,300	1,115	6,427	2,883	2,055	946	676
30–59	5,792	745	2,679	7,296	1,619	−136	1,038	1,092
60–89	2,536	1,108	2,287	2,420	1,569	−56	118	464
90–179	7,011	2,813	2,689	2,083	201	1,015	4	2,406
180–364	2,630	3,096	2,335	−1,393	386	1,125	−117	437
365–999	583	942	1,673	1,340	1,074	816	560	−1,158
All	32,519	11,785	14,621	20,530	9,987	6,457	3,074	4,286

- Fact #2: Net demand positively related to dollar expensiveness.



- Fact #3: Making marking is risky (top panel) and profitable (bottom panel), about $1 mi per year per market maker. Thus, dealers are compensated for the substantial risk they take on.



- Fact #4: Regression evidence of demand pressure and the interaction with the dealer's P&L: effect of demand is stronger after market-maker losses.

Index-option expensiveness explained by end-user demand with control variables

	Non-market-maker demand		
Constant	0.020	0.021	0.044
	(2.75)	(2.79)	(1.61)
Demand	4.23×10^{-5}	4.03×10^{-5}	4.16×10^{-5}
	(4.20)	(3.86)	(4.44)
P&L × Demand		-8.43×10^{-14}	-1.25×10^{-13}
		(−1.22)	(−1.73)
Volatility			-1.42×10^{-1}
			(−0.88)
S&P Return			8.65×10^{-3}
			(0.23)
Adj. R^2 (%)	31.0	31.9	34.3

The SPX excess implied volatility is regressed on non-market-maker demand as well as control variables, 1997/10–2001/12. The demand across contracts is weighted using our model with jump risk. The controls are (i) the product between lagged monthly market-maker profit and demand, (ii) current S&P 500 volatility, and (iii) the lagged monthly S&P 500 return. t-statistics computed using Newey-West are in parentheses. Demand has a positive effect on implied volatility, and the negative coefficient on the interaction between market-maker profits and demand pressure means that the effect of demand is larger following market-maker losses.

4. Volatility and the Real Economy

- Bloom (2009) studies the impact of "uncertainty shocks."
- Motivating figure from before

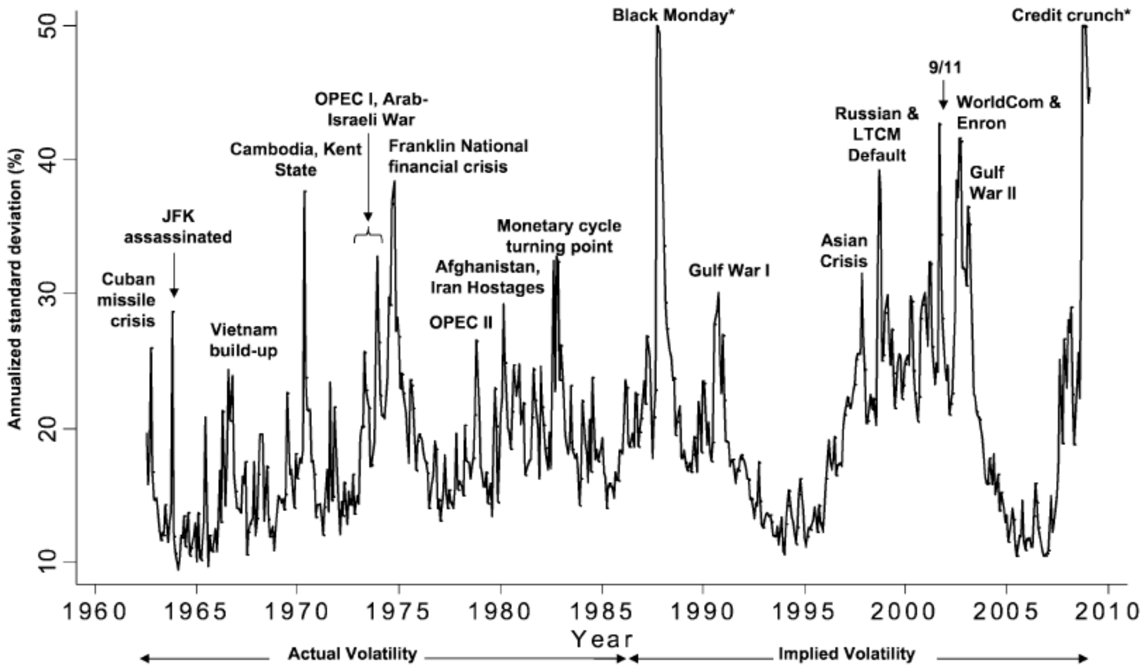


FIGURE 1.—Monthly U.S. stock market volatility. *Notes*: Chicago Board of Options Exchange VXO index of percentage implied volatility, on a hypothetical at the money S&P100 option 30 days to expiration, from 1986 onward. Pre-1986 the VXO index is unavailable, so actual monthly returns volatilities are calculated as the monthly standard deviation of the daily S&P500 index normalized to the same mean and variance as the VXO index when they overlap from 1986 onward. Actual and VXO are correlated at 0.874 over this period. A brief description of the nature and exact timing of every shock is contained in Appendix A. The asterisks indicate that for scaling purposes the monthly VXO was capped at 50. Uncapped values for the Black Monday peak are 58.2 and for the credit crunch peak are 64.4. LTCM is Long Term Capital Management.

- Volatility is endogenous, and Bloom (2009) uses indicator functions corresponding to the peaks in the figure above.

- Bloom (2009) estimates a VAR with monthly data from June 1962 to June 2008.
- The state vector includes
  1. The log stock market index.
  2. The volatility indicator.
  3. The Federal funds rate.
  4. The log of average hourly earnings.
  5. The log of the consumer price index.
  6. Hours.
  7. The log of employment.
  8. The log of industrial production.
- Logic for the ordering:
  - Shocks simultaneously affect asset prices and volatility (series 1-3), then labor/goods prices (series 4-5), and then quantities (series 6-8).
  - By putting the stock market first, we already control for the impact of the stock market when tracing out the impact of volatility shocks.
- Instead of using the level of volatility, Bloom (2009) uses an indicator variable to only use the exogenous nature of these events.
- All series have been de-trended using a Hodrick-Prescott filter (note: the standard HP filter is 2-sided).

- Key figures

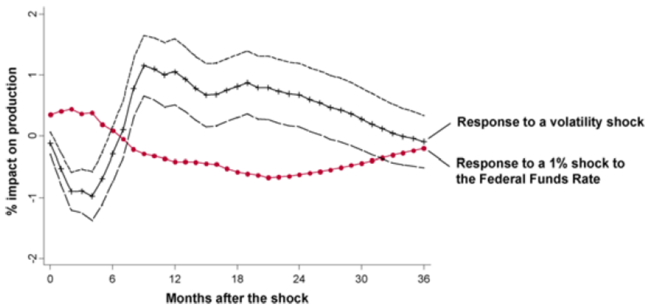


FIGURE 2.—VAR estimation of the impact of a volatility shock on industrial production. *Notes*: Dashed lines are 1 standard-error bands around the response to a volatility shock.

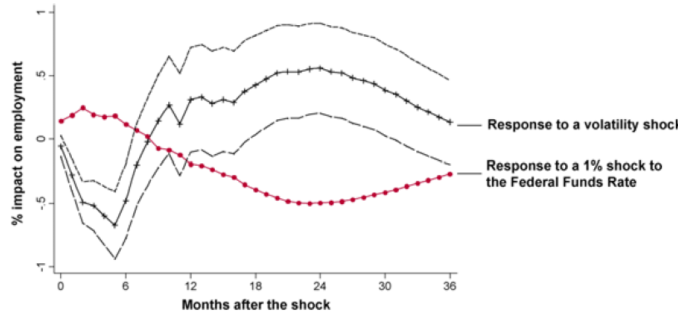


FIGURE 3.—VAR estimation of the impact of a volatility shock on employment. *Notes*: Dashed lines are 1 standard-error bands around the response to a volatility shock.

- Both production and employment fall when uncertainty increases, and then bounce back, but overshoot.
- Bloom (2009) develops a model to reconcile these findings.
- Key innovation: Heteroscedastic aggregate demand and productivity shocks in a firm-level model.
- Intuitively, fixed costs and partial irreversibilities lead to inaction regions. Higher uncertainty increases the inaction region and firms temporarily pause their investment and hiring.
- The model replicates the overshooting in the *level of employment* (not just hiring to catch up to trend).

- Bloom (2009) triggered a large literature trying to understand how to measure uncertainty, what drives uncertainty, and how it affects the real economy.

- One prominent example is Jurado, Ludvigson, and Ng (2015).

- Instead of using stock market volatility, they measure macro uncertainty using a large number of macro-economic time series.

- Consider N_y macro series, $Y_t = (y_{1t}, \ldots, y_{N_y t})'$.

- Define uncertainty for series j, h periods ahead, as

$$U^y_{jt}(h) = \sqrt{\mathbb{E}\left[\left(y_{j,t+h} - \mathbb{E}_t\left[y_{j,t+h}\right]\right)^2\right]},$$

  where this definition uses the *conditional* expectation of $y_{j,t+h}$.

- The aggregate uncertainty measure at time t, h periods ahead, is given by

$$U^y_t(h) = \sum_{j=1}^{N_y} w_j U^y_{jt}(h) \rightarrow_p \mathbb{E}_w\left[U^y_{jt}(h)\right].$$

- High uncertainty means economy has become less predictable.

- Measuring uncertainty requires a model for conditional expectations and a stochastic volatility model for the innovations.

- Measuring uncertainty
  - To proxy for the information available to agents, estimate a factor model for a large number (132) of macro-economic time series, X_t, i.e.,

$$X_{it} = \Lambda_i^{F\prime} F_t + e_{it}^X,$$

    where $dim(F) << dim(X)$.
  - To forecast $y_{j,t+1}$, they use lags of y_{jt}, the factors, and additional variables, W_t

$$y_{j,t+1} = \phi_j^y(L)y_{jt} + \gamma_j^F(L)F_t + \gamma_j^W(L)W_t + v_{j,t+1}^y.$$

  - Stack all factors together (i.e., y_t, F_t, and W_t) and estimate a VAR.
  - Estimate a stochastic volatility model for the residuals, e.g.,

$$\log(\sigma_t^F)^2 = \alpha^F + \beta^F \log(\sigma_{t-1}^F)^2 + \tau^F \eta_t^F, \eta_t^F \sim N(0,1), i.i.d.$$

- Macro uncertainty versus stock market uncertainty

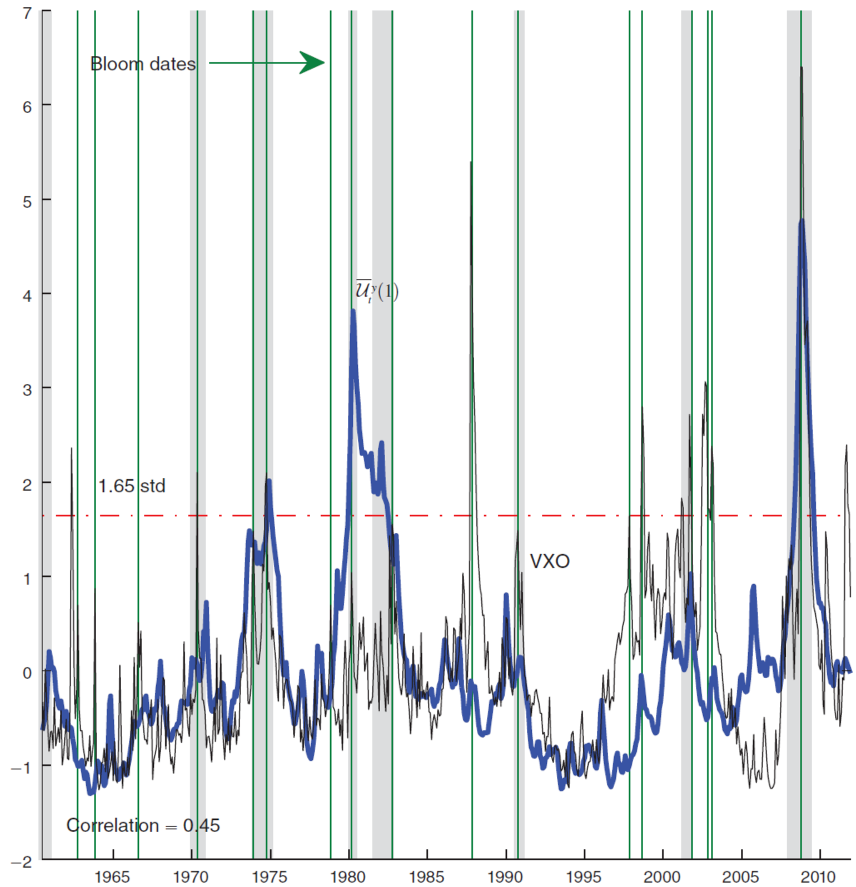


FIGURE 5. STOCK MARKET IMPLIED VOLATILITY AND UNCERTAINTY

Notes: This plot shows $\overline{\mathcal{U}}_t^y(1)$ and the VXO index, expressed in standardized units. The vertical lines correspond to the 17 dates in Bloom (2009) Table A.1, which correspond to dates when the VXO index exceeds 1.65 standard deviations above its HP (Hodrick and Prescott 1997) filtered mean. The horizontal line corresponds to 1.65 standard deviations above the unconditional mean of each series (which has been normalized to zero). The data are monthly and span 1960:7–2011:12.

- Many sharp spikes in stock market volatility do not appear in macro uncertainty. E.g., 1987 was a large spike in stock market volatility but nothing happened macro volatility.

- Re-estimate the same VAR as before, using macro uncertainty

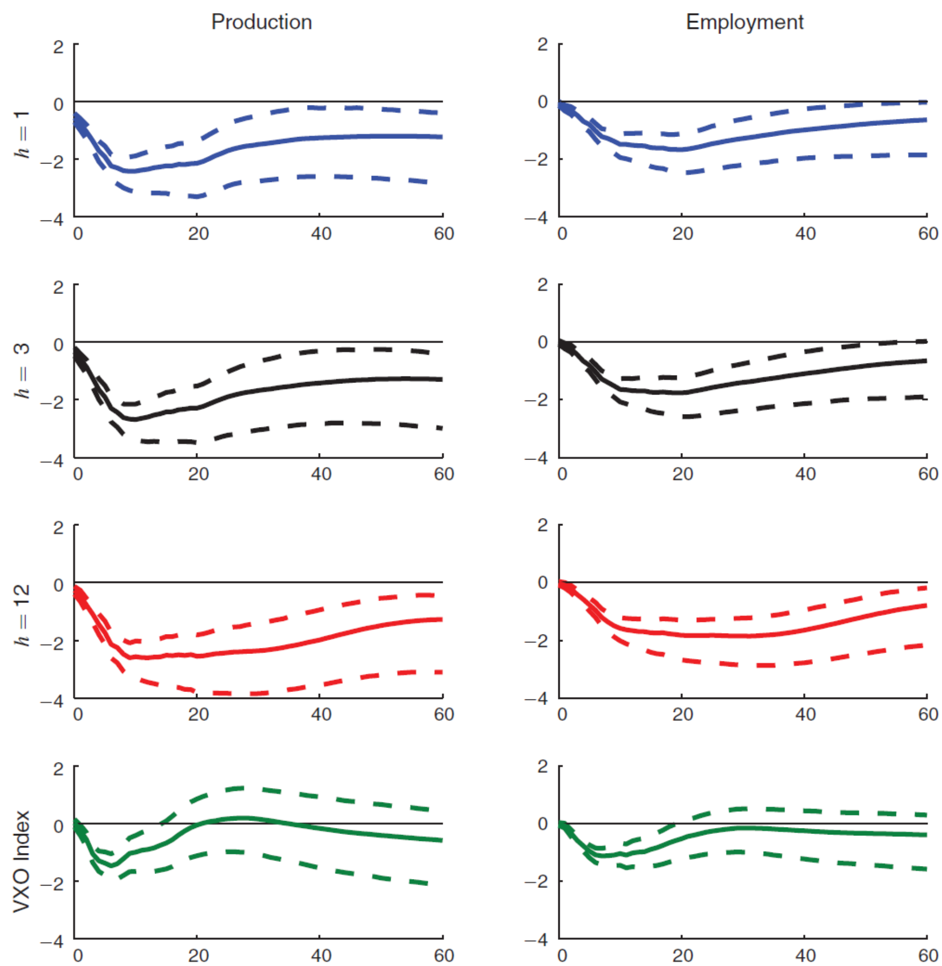
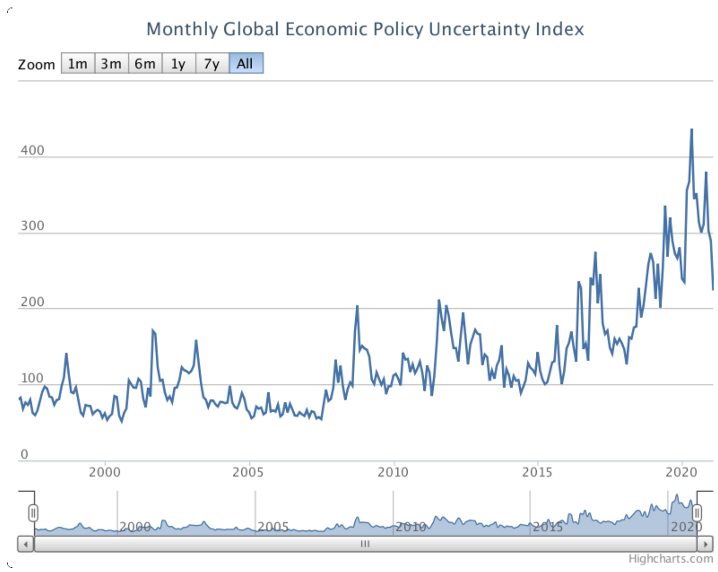


FIGURE 7. IMPULSE RESPONSE OF PRODUCTION AND EMPLOYMENT FROM ESTIMATION OF VAR-8 USING $\overline{\mathcal{U}}_t^y(h)$ OR VXO AS UNCERTAINTY

Notes: Dashed lines show 68 percent standard error bands. The data are monthly and span the period 1960:7–2011:12.

- h measures the uncertainty in months ahead (i.e, one month, one quarter, and one year).
- Important differences:
  - Macro uncertainty has larger and more persistent effects than stock market volatility.
  - No evidence of overshooting.

- Baker, Bloom, and Davis (2016) create an economic policy uncertainty index from news stories (10 large newspapers in the U.S.), federal tax policies that are set to expire in the future, and Survey of Professional Forecasters' dispersion in forecasts about inflation and government spending.



- Available for multiple countries on this web site.
- Used to explain stock market volatility, predict stock returns, predict macro-economic aggregates.