How beliefs respond to news: implications for asset prices

Ian Dew-Becker, Stefano Giglio, and Pooya Molavi*

May 8, 2025

Abstract

This paper studies the implications of a simple theorem, which states that for arbitrary underlying dynamics and diffusive information flows, the cumulants of Bayesian beliefs have a recursive structure: the sensitivity of the mean to news is proportional to the variance; the sensitivity of the nth cumulant to news is proportional to the $n+1$th. The specific application is the US aggregate stock market, because it has a long time series of high-frequency data along with option-implied higher moments. The model qualitatively and quantitatively generates a range of observed features of the data: negative skewness and positive excess kurtosis in stock returns, positive skewness and kurtosis and long memory in volatility, a negative relationship between returns and volatility changes, and predictable variation in the strength of that relationship. Those results have a simple necessary and sufficient condition, which is model-free: beliefs must be negatively skewed in all states of the world.

1 Introduction

Motivation

The US stock market is a compelling laboratory for studying belief dynamics: not only is it deeply important intrinsically, it also is the single richest source of data on expectations. Under very general conditions – not even requiring complete rationality – a security's price is

*Dew-Becker: Federal Reserve Bank of Chicago; Giglio: Yale University; Molavi: Northwestern University. The views in this paper are those of the authors and do not represent those of the Federal Reserve System or Board of Governors. We appreciate helpful comments from seminar participants at the Transatlantic Theory Conference, Saieh Fellows Conference, the Bank of Canada, Duke, the San Francisco and Chicago Feds, Stanford, Indiana, and Yale.

the expected discounted value of its future cash flows. In that sense, the stock market has over a century of data on expectations, in modern times at frequencies that are nearly continuous, and furthermore data on options gives measures of higher-order moments of beliefs.[1] This paper's basic goal is to understand the dynamics of beliefs in a general model of information acquisition, and in particular to understand relationships among the conditional moments.

In the US stock market, there is an extremely strong negative relationship between the aggregate level of prices (e.g. the level of the S&P 500) – again, expectations – and their conditional variance. That negative correlation is known as the **leverage effect**.[2] Many models have been proposed to explain that phenomenon. A very natural hypothesis is that it comes from high volatility raising discount rates (French, Schwert, and Stambaugh (1987)). However, the link between volatility and risk premia is surprisingly tenuous.[3] Instead, we follow a different strand of the literature, focusing purely on belief dynamics.[4] Part of our aim is to understand whether there is a necessary and sufficient condition for belief dynamics to be associated with a leverage effect.

Beyond that first relationship, there are numerous other features of returns to understand – negative skewness and positive excess kurtosis in returns, positive skewness and excess kurtosis along with long memory in the volatility of returns, and a strong relationship between conditional skewness and the future covariance between prices and volatility. No other dataset provides the same detail and breadth of moments that can be used for testing a model of beliefs, and there is no unified framework that quantitatively captures all of those features of the data.

Contribution

The basic structure of the paper is to generate predictions from a very general model of belief formation and then examine them quantitatively in stock market data.

The theoretical structure is built around the idea that agents fundamentally want to know the discounted value of a security's cash flows. There are many ways that cash flows and information can be modeled, but they are all broadly asking how expectations are updated as information arrives. We therefore study a simple but general setup: the net present value

[1] The VIX volatility index (based on the so-called model-free implied variance of Britten-Jones and Neuberger (2000)) is the most well-known option-implied moment. Work on option-implied distributions goes back to Breeden and Litzenberger (1978).

[2] See Merton (1980) and French, Schwert, and Stambaugh (1987), among many, many others.

[3] See Lettau and Ludvigson (2010) for a review. Moreira and Muir (2017) show how an investor historically could have taken advantage of this fact.

[4] Among others, see David (1997), Veronesi (1999), Weitzman (2007), David and Veronesi (2013), Collin-Dufresne, Johannes, and Lochstoer (2016), Johannes, Lochstoer, and Mou (2016), Kozlowski, Veldkamp, and Venkateswaran (2018), Farmer, Nakamura, and Steinsson (2024), Wachter and Zhu (2023), and Orlik and Veldkamp (2024). While those papers aassumem rationality, there is also a large behavioral literature that focuses on belief dynamics, e.g. Gennaioli, Schleifer, and Vishny (2015).

(NPV) follows some arbitrary process, and agents continuously receive signals about it. The signals represent in reduced form the aggregate of all the information people observe in reality. Since the NPV process is essentially unconstrained, the analysis nests a wide range of specifications that have been studied in the literature.

The results are stated in terms of the cumulants of agents' conditional distributions for the NPV (which we refer to as fundamentals). Recall that the first three cumulants of a distribution are equal to the first three central moments (in the specific results, we do not go past the third cumulant). The paper's main theoretical result is a transparent recursive relationship among the cumulants. Specifically, the sensitivity of the first moment (which is the price agents will pay for the security) to news is proportional to the second moment, and in fact the sensitivity of the nth cumulant to news is proportional to the $n+1$th cumulant.

The paper's core motivating fact is that aggregate stock returns are negatively correlated with innovations to volatility. When agents are learning about fundamentals, that happens if and only if agents' conditional third moment for fundamentals is negative, and in fact the magnitude of the return-volatility relationship is proportional to the conditional third moment, representing a strong and testable prediction.

Beyond sensitivity, the theorem also yields expressions for the drift in the cumulants. Interestingly, mean-reversion in volatility is generically nonlinear: instead of depending on the current level of volatility, the drift in volatility is proportional to its own current square. That fact yields the sort of long memory that has been observed in stock market volatility. Long memory is therefore a nearly inevitable feature of information acquisition, appearing regardless of dynamics of fundamentals (except the knife-edge case of a linear-Gaussian process).[5]

Long memory represents a prediction that is distinct from what would be implied by a model in which the leverage effect is driven purely by risk premia. A second prediction that also distinguishes the model is that the magnitude of the leverage effect (the coefficient in a regression of volatility changes on market returns) should be equal to the current conditional skewness of returns divided by 3. The fact essentially follows from the model's prediction that prices do not jump and that innovations to volatility are perfectly spanned by innovations to prices. In the data, the leverage effect coefficient lines up strikingly well with the model's prediction, which implies that while obviously prices do jump and there

[5]There is a long-running literature on long memory in stock market volatility going back to Mandelbrot (1963). It is sometimes modeled as being fractionally integrated, and such processes can be constructed as the sum of an infinite number of independent components (e.g. Granger (1980) and Mandelbrot, Calvet, and Fisher (1997)). A simple univariate example of nonlinear decay in uncertainty is to note that with a Gaussian prior with variance σ_0^2 and Gaussian signals with variance σ_S^2, the posterior variance after obsering N signals is $\left(\sigma_0^{-2} + N\sigma_S^{-2}\right)^{-1}$, which shrinks polynomially rather than geometrically.

is unspanned variation, the model's core mechanism is the driving force underlying stock market return skewness.

After developing a few more theoretical results, we move on to a quantitative analysis of the model's predictions. First, we examine a very simple parametric model with three free parameters that has the characteristics we find are necessary for matching the data. That model (with a constant probability of exponential jumps) is able to match the first four moments of returns, volatility, and changes in volatility, along with volatility's autocorrelations and its relationship with returns. The results have two implications. First, they show the mechanism is quantitatively relevant. Second, the addition of an extremely simple learning process generate quantitatively significant nonlinearity – enough to match what is observed in stock return dynamics.[6]

Second, and perhaps more relevantly, we show how to derive nonparametric predictions from the model – tests and estimates that can be obtained without knowing anything about the underlying dynamic process for fundamentals. First, the model has predictions for the relationship between volatility, its own lag, returns, and option-implied skewness, that we test and find hold well in the data. Second, it is possible to estimate agents' implied uncertainty about the level of fundamentals without knowing the underlying model. In US stock market data, we estimate that uncertainty to have a standard deviation of between 10.4 and 16.5 percent. In a survey administered by Yale University since the 1980's, cross-sectional disagreement about the fundamental value of the stock market has a standard deviation of 17.0 percent, which provides some independent support for our estimate (subject to the usual caveat that disagreement and uncertainty are theoretically distinct).

Related work

As discussed above, this paper is most closely related to past work studying non-Gaussian filtering problems, including Veronesi (1999), David and Veronesi (2013), and Kozlowski, Veldkamp, and Venkateswaran (2018), among others. The first two papers study learning about states, while the last is about learning about time-invariant parameters, but both types of learning are accommodated within our setup. A general feature of non-Gaussian learning is that is not very tractable – solutions are often characterized in terms of some differential equation that must be solved numerically. This paper makes some progress on that front because it is able to directly describe the dynamics of key features of interest – agents' conditional moments – in a general setting.[7]

[6]Again, past work noted above has shown that learning can help explain stock return dynamics. The point here is that these results are extremely general and robust and not dependent on the specific settings studied in past work.

[7]See also Abel, Eberly, and Panageas (2013) and Collin-Dufresne, Johannes, and Lochstoer (2016) , among many others.

Additionally, an important distinction from past work is that even though we assume agents know the structure of the economy (which is not to say that they know the value of the parameters, just that they know the structure and what parameters they need to estimate), it is still possible to obtain testable predictions even if the econometrician does not know the true underlying structure, and in fact it is even possible to estimate agents' conditional uncertainty about fundamentals.

As noted above, the force in the model that generates most of the core predictions is that agents have negatively skewed beliefs about fundamentals. In the context of the model, that happens because fundamentals themselves follow a skewed process. This paper does not explain where that fundamental skewness, it just examines its implications. Skewness in fundamentals, though, is observed in many studies of the real economy (e.g. Sichel (1993), McKay and Reis (2008), Morley and Piger (2012), Berger, Dew-Becker and Giglio (2020), and Dupraz, Nakamura and Steinsson (n.d.), among others), and there is theoretical work that tries to explain it (e.g. Ilut et al. (2018), Straub and Ulbricht (n.d.), Kozeniauskas et al. (2018), Gilchrist and Williams (2000), Kocherlakota (2000), Hansen and Prescott (2005), Bianchi (2011), and Bianchi et al. (2017)).

The quantitative example the paper studies is closely related to work on rare disasters (e.g. Rietz (1988) and Barro (2006)). Even though the probability of a disaster (here just a jump in the fundamental value of stocks) is constant, at any given time agents are unsure whether a disaster has occurred, so their subjective distribution over future returns varies over time as though the probability of a disaster is time-varying, as in Gabaix (2012) and Wachter (2013) (see also Wachter and Zhu (2023) for a model of learning with rare disasters). Baker et al. (2025) provide evidence on what the actual events are that cause jumps in stock prices.

Finally, Altig et al. (2022) and Bachmann et al. (2024) are important precursors to this work for studying, in the context of a survey, not just the level of managers' uncertainty but how it is updated over time.

Outline

The remainder of the paper is organized as follows. Section 3 describes the model structure and gives the main theoretical result. Section 4 then examines the theoretical predictions and section 5 studies some extensions and robustness to certain assumptions. Last, sections 6 and 7 take the model to the data, studying both a calibration and nonparametric tests of the theory, and section 8 concludes.

2 Motivating facts

We begin with a set of motivating facts. All of the conditional moments in this section are measured by forecasting realized moments with option-implied moments. For example, we estimate the conditional volatility of returns by projecting realized volatility onto current option-implied volatility (the VIX index). The results are highly similar just using the option-implied moments themselves.[8]

Table 1: Stock market return and volatility moments

	Stock market	—— Volatility ——-	
Moment	Daily return	Level	Daily change
Std. dev.	1.14	6.64	1.39
Skewness	-0.26	2.19	1.51
Kurtosis	12.68	11.57	30.03
Corr. w/ R_t			-0.78

Note: The table reports empirical moments of stock market returns and volatility (level and daily changes). Volatility is the fitted value of a projection of realized volatility onto the vix.

Table 1 reports the second, third, and fourth central moments for daily market returns, their conditional volatility (in annualized standard deviation units), and the daily change in that conditional volatility. Daily market returns are slightly negatively skewed, while volatility is highly positively skewed in both levels and changes. All three series have severe excess kurtosis, consistent with time-variation in their volatilities or the presence of large jumps.

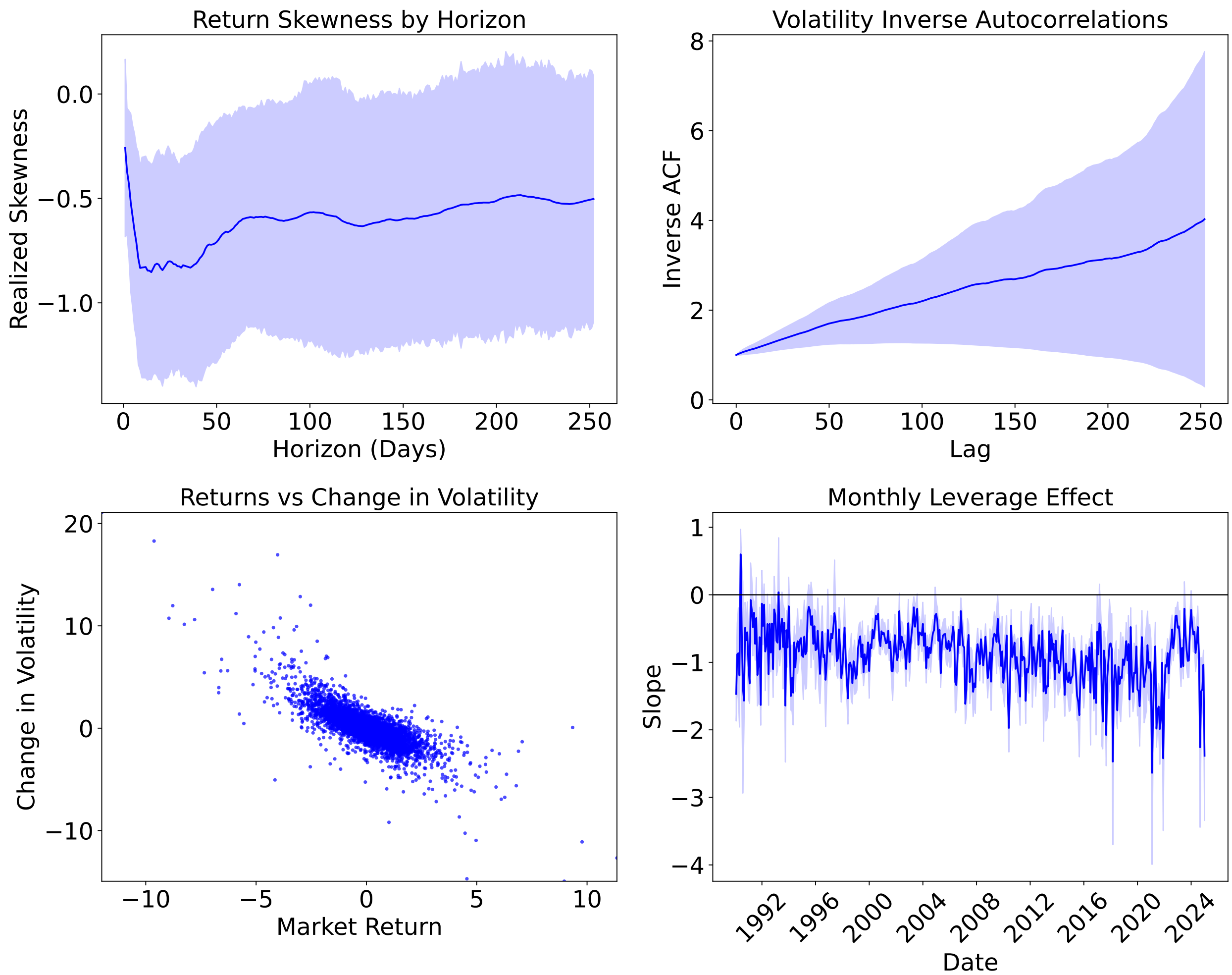
The top-left panel of figure 1 plots the realized skewness of returns at horizons of 1 to 252 days. Skewness becomes significantly more negative as the horizon initially increases, and then reverts somewhat back towards zero, reaching about 0.8 for annual returns.

The bottom-left panel of figure 1 is a scatter plot of daily market returns against the daily change in the volatility, showing the strong negative correlation. The correlation coefficient in the sample is -0.70. The bottom-right panel plots estimates of the regression coefficient in every month between 1990 and 2025, showing that the negative relationship is not isolated to particular episodes – it holds in every month in the sample except for two.

Finally, the top-right panel of figure 1 plots the inverse autocorrelations of volatility – i.e. the jth point on the x-axis is $1/corr\left(vol_t, vol_{t-j}\right)$. The reason to plot the inverse

[8] For volatility, projecting realized volatility on the VIX yields a series that has a level shift, due to the variance risk premium, but nearly the same standard deviation. The adjustment to using a projection has a bigger impact on conditional skewness, studied below, as is important for properly linking the model with the data.

Figure 1: Motivating evidence



Note: Top left: realized skewness of returns computed over different holding periods. Top right: inverse of autocorrelations of volatility at different lags (in days). Bottom left: scatterplot of daily changes in volatility against daily market returns. Bottom right: monthly series of the slope of a regression (using daily data within each month) of change in volatility onto returns. All figures use daily data from 1990. Shaded areas are 95% confidence intervals.

autocorrelations is to help emphasize the deviation from the exponential decay that would be expected if volatility followed an ARMA process. The fact that $corr\left(vol_t, vol_{t-j}\right)^{-1}$ grows linearly is consistent with polynomial decay in the autocorrelations, as in fractionally integrated models such as Ding, Granger, and Engle (1993) and Bollerslev and Mikkelsen (1996) (among many others).[9]

The theoretical analysis that follows will show how a simple and general model of information acquisition can qualitatively generate all of the moments reported in this section. Section 6 shows that the mechanism is also quantitatively realistic.

3 Model setup and solution

This section describes the asset pricing framework we study. It is designed to lead to a standard filtering problem, which we then show can be tractably analyzed.

3.1 Model setup

3.1.1 Dynamics of fundamentals

Stocks pay some cash-flow D_t and that there is a stochastic discount factor M_t such that, conditional on any information set $\mathcal{I}_t$, prices satisfy

$$P_t\left(\mathcal{I}_t\right) = E\left[\int_{s=0}^{\infty} \frac{D_{t+s}M_{t+s}}{M_t} ds \mid \mathcal{I}_t\right] \tag{1}$$

where the notation $P_t\left(\mathcal{I}_t\right)$ emphasizes the dependence of prices on the information set. This specification nests a wide range of models – the stochastic discount factor encodes risk aversion and any other drivers of state prices and can, under certain assumptions, also represent distortions in beliefs.[10] All the results can easily be mapped into settings where state prices are not at issue, such as surveys of expectations, by simply setting $M_t = 1\ \forall\ t$.

At any given time, the full set of information that an agent could possibly know is represented by θ_t. The fundamental value of the asset is its price conditional on complete knowledge of θ_t,

$$X_t \equiv P_t\left(\theta_t\right) \tag{2}$$

[9]Though note that fractional integration is an asymptotic concept and there is an ARMA model that can match any finite set of autocorrelations.

[10]It is worth noting here that we are taking the SDF as given. While that will be valid in the context of the paper's analysis, more generally it is natural to think that it in fact might change if agents' information set changes.

An extreme case is perfect foresight, in which θ_t contains complete knowledge of all values of cash flows in the future. But θ_t can also be coarser.

The model is driven by the dynamics of the state variable θ_t and the function P_t mapping from information to asset prices. Assumptions 1–3 in the appendix give the required technical restrictions on them. The conditions required for the paper's results to hold are just those that make filtering possible. Essentially the first and second moments of X_t (and its Fourier transform) need to exist and not be too pathological. The complex exponential of X_t must have a conditional mean process that is not too exotic and X_t itself must have a finite second moment. The jump diffusions and continuous-time ARMA processes, both in scalar and vector forms, that are typically studied in economics will be acceptable here. As long as filtering is possible, the paper's results hold.

Finally, note that one particular case for θ_t is that it is the outcome of a learning model. For example, cash-flows might have a mean growth rate that is unknown, so that expectations depend on the average growth rate observed up to date t.[11]

3.1.2 Information flows

Agents observe a history of signals denoted by Y^t. If the payoff-relevant information in Y^t is a subset of that in θ_t (i.e. $P_t(Y^t, \theta_t) = P_t(\theta_t)$),[12] then by the law of iterated expectations,

$$P_t\left(Y^t\right) = E\left[X_t \mid Y^t\right] \tag{3}$$

(3) is a standard filtering problem. In principle all that is left is to specify Y. Prices are a simple expectation here because X itself already includes risk adjustments represented in state prices via the M process.

The one final wrinkle is that because aggregate stock prices display trend growth, they are typically modeled in logs. If X is an arithmetic process (which it might be in the case of inflation or interest rates), we could directly apply (3). To analyze stock prices, which are typically modeled as a geometric process, we restate the filtering problem in logs,

$$\begin{aligned} p_t &\equiv E\left[x_t \mid Y^t\right] && (4)\\ \text{where } x_t &= \log X_t && (5) \end{aligned}$$

[11]Specifically, if the dividend process satisfies $dD_t = \mu_D dt + \sigma_D dB_t$ for a Brownian motion B, then $E_t[D_{t+s}] = D_t + s\frac{D_t}{t}$. D_t then remains a Markov process with respect to its own natural filtration.

[12]More formally, conditional on θ_t, $\int_{s=0}^{\infty} \frac{D_{t+s}M_{t+s}}{M_t} ds$ is independent of Y^t. If an agent happened to know the true state θ_t, the signal history Y^t would contain no additional information.

and p_t is the **log** of the price agents actually pay, given Y^t. That is the single approximation step in the analysis. Transformations like this are not uncommon – analyses of macro-finance models very often rely on the Campbell–Shiller approximation, for example, and an alternative would be to motivate (4) that way. Clearly by treating log prices as the expectation of log fundamentals, a form of (relatively mild) risk aversion appears here.[13]

Last, we must specify how information arrives. We assume that all of the agents' information is generated via the process

$$dY_t = x_t dt + \sigma_{Y,t} dW_t \tag{6}$$

where $\sigma_{Y,t}$ follows some exogenous process (subject to assumption 3 in appendix A.1.1.1) and the information set Y^t represents the history of the Y process up to date t.[14] A reasonable benchmark is that $\sigma_{Y,t}$ is constant, but it could also vary over time, giving a form of time-varying uncertainty. Furthermore, full information is nested as $\sigma_{Y,t} \to 0$.[15]

Obviously this is not the only possible information structure. Agents could receive signals about nonlinear functions of x_t, such as its moments, or about θ_t, which might contain relevant information about the future path of x. Additionally, they might draw inferences about θ_t from realized cash-flows.[16] The Y process can be thought of as a simplification that captures all the information agents receive in a single factor. And given that x represents how agents would value stocks if they had complete information, it makes a certain amount of sense to assume that it is what agents learn about. The assumption that information flows diffusively also matters for the analysis, but it is not completely restrictive – section 5.2 discusses how the results apply when when there are discrete information revelation events. The analysis is extremely general in the dynamics for fundamentals, represented by x, but pays for that generality with a somewhat tight restriction on the information structure.

The structure here is motivated by an asset pricing problem, but it is a much more general setup. x is just some latent object of interest – it could be trend inflation, for example. Then $E[x_t \mid Y^t]$ would represent agents' expectations of trend inflation given their history of signals.

[13] We directly specify equation (4), essentially as a behavioral assumption, rather than starting from the Campbell–Shiller approximation, precisely because it is transparent about exactly what the elision is here.

[14] Formally, $\sigma_{Y,t}$ is measurable with respect to the filtration induced by Y.

[15] Though note that while $P(Y^t)$ converges pointwise to $P(\theta_t)$ as $\sigma_Y \to 0$, it need not converge uniformly (in particular if X_t has jumps).

[16] Note that in the case of US stocks, cash-flows are strictly pre-determined. Dividends, for example, are announced well in advance of their payment.

3.2 Solution to the filtering problem

The paper's results primarily involve the dynamics of the first two moments of agents' posteriors, denoted here by $\kappa_{1,t}$ and $\kappa_{2,t}$ (with $\kappa_{3,t}$ naturally being the third moment).

Proposition 1 *The posterior mean ($\kappa_{1,t}$) and variance ($\kappa_{2,t}$) satisfy*

$$
\begin{aligned}
dp_t &= d\kappa_{1,t} = \frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[dx_t\right] && (7)\\
d\,\mathrm{var}_t\left[x_t\right] &= d\kappa_{2,t} = \frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) - \frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt && (8)\\
&\quad + E_t\left[d\left[x\right]_t\right] + 2\,\mathrm{cov}_t\left(x_t, dx_t\right) && (9)
\end{aligned}
$$

where $[x]$ is the total quadratic variation process of x and $E_t\left[\cdot\right] \equiv E\left[\cdot \mid Y^t\right]$.

For the mean, $\kappa_{1,t}$, the first term says that the sensitivity to news is equal to current uncertainty ($\kappa_{2,t}$) multiplied by the precision of the signal, while the second term is simply the current expected drift. The intuition for the gain is simple: $\kappa_{2,t}/\sigma_{Y,t}^2 = \mathrm{cov}_t\left(x_t, dY_t\right)/\mathrm{var}_t\left(dY_t\right)$ is a local regression coefficient.

The dynamics of the conditional variance are similar. The gain is now the third moment times the precision of the signal. That is again because $\kappa_{3,t} = E\left[\left(x_t - E_t\left[x_t\right]\right)^3\right]$ is equal to $\mathrm{cov}_t\left(\left(x_t - E_t\left[x_t\right]\right)^2, dY_t\right)/dt$, so $\kappa_{3,t}/\sigma_{Y,t}^2$ is the local regression coefficient. We discuss the drift term $\frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt$ further below. $E_t\left[d\left[x\right]_t\right]$ represents the expected accumulation of variance in x (i.e. due to shocks), and $2\,\mathrm{cov}_t\left(x_t, dx_t\right)$ is the accumulation of uncertainty due to x effectively "spreading out" over time.

Remark 1 *Prices follow an Itô diffusion and satisfy*

$$
\begin{aligned}
dp_t &= \mu_{p,t}dt + \sigma_{p,t}d\tilde{W}_t && (10)\\
\textit{where } d\tilde{W}_t &\equiv \sigma_{Y,t}^{-1}\left(dY_t - E_t\left[x_t\right]dt\right) \;\; \textit{is a Brownian increment,}\\
\mu_{p,t}dt &= E_t\left[dx_t\right], \; \textit{and } \sigma_{p,t} = \kappa_{2,t}/\sigma_{Y,t}
\end{aligned}
$$

Prices follow a completely standard continuous diffusive process. What the model delivers is simply a very specific structure for the conditional volatility process. All of the predictions for prices ultimately follow from the dynamics of volatility, which is itself a diffusion driven by the same Brownian motion W (see corollary 2 below).

3.2.1 General result

While those equations will be enough for the present paper, they suggest a broader result: the gain coefficients seem to satisfy a recursion. That recursion turns out to hold for the cumulants of agents' posteriors, which are the derivatives of the log characteristic function (properly factoring out powers of $\sqrt{-1}$). The first three cumulants are equal to the first three central moments, which is why we did not need to mention them yet.

Denote the n-th cumulant of the time-t conditional distribution of x_t by $\kappa_{n,t}$.

Theorem 1 *Given (6) and restrictions on x_t given in appendix A.1.1, for all n for which the $n+1$th cumulant exists*[17]

$$\begin{aligned} d\kappa_{k,t} &= \frac{\kappa_{k+1,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t[x_t]dt\right) - \frac{1}{2\sigma_{Y,t}^2}\sum_{j=2}^{k}\binom{k}{j-1}\kappa_{j,t}\kappa_{k-j+2,t}dt \\ &\quad + \sum_{j=1}^{k}\binom{k}{j} B_{k-j}\left(-\kappa_{1,t},\ldots,-\kappa_{k-j,t}\right)\mathbb{E}_t[d(x_t^j)] \end{aligned} \tag{11}$$

where B_j denotes the jth complete exponential Bell polynomial.

That result follows from a straightforward application of textbook results in Lipster and Shiryaev (2013) and Bain and Crisan (2009).[18] Theorem 1 shows that the recursion for the gain carries through to all the cumulants – the gain of the nth cumulant is the $(n+1)$th cumulant times the precision of the signal.

4 Predictions

We now examine the predictions of proposition 1 for the behavior of returns. Many of the results pertain to return volatility, which we define as the instantaneous volatility process

[17]Since the cumulants are derivatives of a function, if $\kappa_{n+1,t}$ exists then all lower-order cumulants also exist. Note that the distribution of x_t conditional on Y^t is necessarily subgaussian, meaning that all moments and cumulants exist (Guo et al. (2011)). So the restriction to n such that the $n+1$th cumulant exists may possibly be satisfied for all n for all processes, but we have not been able to verify that.

[18]Theorem 1 is closely related to results in Dytso, Poor, and Shamai (2022), with two key differences. First, x_t here is dynamic instead of constant. Second, theorem 1 enables the calculation of the evolution of the conditional cumulants from knowledge only of the priors. Surprisingly, as Dytso, Poor, and Shamai (2022) discuss, there do not appear to be any other earlier precedents to the family of results in their work and ours.

for log prices[19]

$$vol_t \equiv \left(\lim_{\Delta t \downarrow 0} E\left[\left(p_{t+\Delta t} - E_t p_{t+\Delta t} \right)^2 \right] / \Delta t \right)^{1/2} \tag{12}$$

$$= \kappa_{2,t} / \sigma_{Y,t} \tag{13}$$

That is, vol_t is simply the diffusive volatility of prices from the representation (10). Again, the conditional volatility of prices depends on agents' current posterior variance over fundamentals, $\kappa_{2,t}$. So, up to $\sigma_{Y,t}$, price volatility measures uncertainty.

Combining equations (8) and (7) from proposition 1, we have

Corollary 2 *When $\sigma_{Y,t}$ is constant, vol_t follows a diffusion satisfying*

$$d\left(vol_t\right) = \underbrace{\sigma_{Y,t}^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}} \left(dp_t - E_t dp_t \right)}_{ⓐ} - \underbrace{\frac{vol_t^2}{\sigma_{Y,t}} dt}_{ⓑ} + E_t \left[d\left[x\right]_t \right] + 2 \operatorname{cov}_t \left(x_t, dx_t \right) \tag{14}$$

The main predictions arise out of the terms ⓐ and ⓑ. ⓐ determines the joint behavior of returns and their higher moments, while ⓑ generates nonlinearity in the dynamics of volatility. The third term is, again, the spreading out of x_t itself, which is driven by the dynamics of fundamentals, rather than the learning that is our focus.[20] Allowing $\sigma_{Y,t}$ to vary is straightforward, just adding an extra term, but we leave it out here because it plays no role in the analysis below – we effectivaly treat it as constant from here forward.

A first point to note is that volatility is generically time-varying. It is only when the model is fully linear and Gaussian that the volatility of prices converges to a constant. If any of the higher-order cumulants is nonzero, that effectively immediately creates a change in volatility. Time-varying volatility by itself is enough to generate, qualitatively, the excess kurtosis observed in stock market returns in table 1. The following sections show how the model can qualitatively generate the broader behavior documented in table 1 and figure 1.

[19] For stocks at high frequency, cash flows are predetermined, and in any case the variance of changes in cash flows for the aggregate US stock market at even the monthly frequency is insignificant compared to changes in prices. The historical variance of monthly returns 2.85×10^{-3}, while the variance of dividend growth is over 600 times smaller – 4.46×10^{-6}. We therefore treat return volatility as equal to price volatility.

[20] In principle, $\operatorname{cov}_t(x_t, dx_t)$ is related to $\kappa_{2,t}$, so it is not completely driven by fundamentals alone. However, the paper's focus will be on the case where fundamentals are a martingale, so that conditional expectations of dx_t are always equal to zero, which also makes the covariance zero.

4.1 Volatility and the leverage effect

Using term ⓐ from equation (14), we get the following result for the leverage effect:

Proposition 2 *The instantaneous coefficient in a regression of changes in the conditional variance of returns on price changes is*

$$\frac{\operatorname{cov}(dp_t, dvol_t)}{\operatorname{var}(dp_t)} = \frac{\kappa_{3,t}}{\sigma_{Y,t}\kappa_{2,t}} \tag{15}$$

The term ⓐ in (14)shows that the presence of a **leverage effect** – the negative correlation between changes in volatility and prices in table 1 and the bottom panels of figure 1 – is completely determined by the third moment of agents' conditional distribution and the noise in agents' signals. The necessary and sufficient condition for the existence of a leverage effect is that $\kappa_{3,t} < 0$: there is a leverage effect if and only if agents' posterior distribution for fundamentals is negatively skewed. And the fact that we observe a leverage effect in the aggregate US stock market in nearly all months in the data, including during severe downturns, then implies that the conditional skewness is negative in essentially all states of the world observed in our sample. The relationship has additionally strengthened over time, which would be consistent with a decline in $\kappa_{3,t}$. We return to this point further in section 7.

The intuition for proposition 2 is straightforward: a negative third moment means that the right tail of the conditional distribution is shorter than the left. When agents receive good news about fundamentals, that tells them they are likely on the narrower side of the distribution, and their conditional uncertainty falls.

4.2 Slow decay in volatility

The term ⓑ in (14) shows how volatility decays. When volatility is high, $vol_t^2 \sigma_{Y,t}^{-1} dt$ also grows, pulling volatility back down towards its steady state. Interestingly, though, unlike standard time-series models (e.g. an AR(1) or Ornstein-Uhlenbeck process), the mean reversion is quadratic, so that the rate of mean reversion rises more than proportionately with increases in volatility.

There is a large empirical literature studying nonlinearity in volatility dynamics in securities markets. The form of mean reversion here is consistent with that literature, in that the decay is non-exponential.[21] When jumps up in vol_t are large relative to its steady-state

[21] See Corsi (2009) for a discussion of some of the evidence (going back at least to Ding, Granger, and Engle (1993)) along with the fact that the data is generally consistent both with strict long memory and also processes that simply approximate it, since formally long memory is defined asymptotically

value, its decay after a time Δt is approximately of the form $1/\left(a+\Delta t\right)$ for a coefficient a that depends on the other parameters of the model. That is exactly the polynomial decay studied in the literature on long memory in volatility, and it is the inverse linear decay that is also observed in the top-right panel of figure 1.

Intuitively, volatility decays nonlinearly because the degree to which agents respond to signals (i.e. the magnitude of the gain) is increasing in uncertainty. When uncertainty is high, agents update strongly in response to signals and learn quickly. As uncertainty falls, agents update less strongly and learning slows. While that phenomenon is well known in linear Gaussian filtering problems (it is a core feature of the Kalman filter), equation (8) shows it is actually a general feature of filtering problems.

The end result is that when investors are learning about fundamentals dynamically, long memory is generic, and only disappears in the knife-edge case of a fully linear Gaussian model, where all higher moments are equal to zero at all times. We examine the model's ability to fit detailed data on volatility dynamics in more detail in section 7.

4.3 Skewness in returns

Since the price process, p, is a diffusion, its instantaneous skewness is not well defined formally. Skewness arises as returns interact with changes in volatility. We get the following result using a second-order approximation for the third moment of returns, defining skewness as usual as

$$skew_t\left(p_{t+\Delta t}\right) \equiv \frac{E_t\left[\left(p_{t+\Delta t}-E_t p_{t+\Delta t}\right)^3\right]}{E_t\left[\left(p_{t+\Delta t}-E_t p_{t+\Delta t}\right)^2\right]^{3/2}} \tag{16}$$

Proposition 3 *The local skewness of returns is*

$$\lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right)\left(\Delta t\right)^{-1/2} = 3\frac{\kappa_{3,t}}{\kappa_{2,t}}\sigma_{Y,t}^{-1} \tag{17}$$

That is, the conditional "instantaneous" skewness of returns again depends on the second and third moments of the posterior. As $\Delta t \to 0$, skewness goes to zero – that is the usual result that returns are locally normal. For small but nonzero values of Δt, skewness is approximately $skew_t\left(x_t\right)\kappa_{2,t}^{1/2}\sigma_{Y,t}^{-1}\left(\Delta t\right)^{1/2}$. That fact provides a link between indexes of the conditional skewness of returns, such as Cboe's option-implied skewness, and the conditional skewness of fundamentals, $skew_t\left(x_t\right)$, which determines the leverage effect:

Corollary 3 *The leverage effect coefficient as decline in proposition 2 is related to the* *skewness of returns via*

$$\lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right) \frac{1}{3} \left(\Delta t\right)^{-1/2} = \frac{\operatorname{cov}\left(dp_t, dvol_t\right)}{\operatorname{var}\left(dp_t\right)} \tag{18}$$

Finally, with respect to the data, this section shows that the model is able to qualitatively match the return skewness documented in table 1 and the top-left panel of figure 1 again as long as $\kappa_{3,t}$ is generally negative. Skewness arises here again due to the term @ in equation 14. When $\kappa_{3,t} < 0$, declines in prices raise volatility, leading to a relatively long tail in returns. Neuberger (2012) and Neuberger and Payne (2021) discuss this mechanism extensively.

4.4 Skewness in volatility

Table 1 shows that both the level and daily changes in stock market volatility are also skewed. The source of that effect is visible if we combine equations (8) and (17) to obtain

$$std\left(vol_t\right) = \frac{1}{3} vol_t \left| \lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right) \left(\Delta t\right)^{-1/2} \right| + o\left(\Delta t^{1/2}\right) \tag{19}$$

Holding the conditional skewness of returns fixed, the volatility of innovations to vol_t scales with vol_t itself. When vol_t falls towards zero, the volatility of its innovations quickly becomes much smaller, while they grow when vol_t rises. That effect creates a long right tail in the level of vol_t. Past work (e.g. Bollerslev, Tauchen, and Zhou (2009)) has emphasized the importance of time-varying vol-of-vol. This present model gets it through an endogenous mechanism. Note also that this variation does not just come from the volatility of fundamentals following a nonlinear process, as in Cox, Ingersoll, and Ross (1985). Finally, as with returns, time-varying volatility in volatility mechanically also generates excess kurtosis in the unconditional distribution of volatility.

4.5 Summary

To briefly summarize the results so far, simple filtering predicts a leverage effect if and only if $\kappa_{3,t} < 0$, and more generally for $\kappa_{3,t} \neq 0$ it generates long memory in volatility, skewness and excess kurtosis in returns, skewness in both levels and changes in volatility, and time-varying volatility of volatility, leading also to excess kurtosis in volatility itself. None of this requires anything more complicated than Bayesian updating in the presence of nonzero higher moments, and the key mechanism lies essentially entirely in equation (14).

4.6 Examples

This section briefly considers two simple examples. Section 6 studies in much more depth a quantitatively realistic example.

4.6.1 Linear Gaussian process

If fundamentals, x, follow a linear Gaussian process then the model's solution is the Kalman filter. p_t is a linear function of the history of signals; its gain and hence conditional variance eventually converges to a constant; and its conditional skewness and all higher cumulants are always equal to zero. There is then no leverage effect, volatility of volatility, or skewness in expectations or volatility.

4.6.2 Markov switching process

Veronesi (1999) studies a two-state switching model in which the latent state x switches between a low and a high value at rates λ_{HL} and λ_{LH}, respectively, and agents have a Gaussian signal as required in theorem 1. In this case, the low and high values of x_t can be normalized to 0 and 1 without loss of generality.

Agents' posterior at any given time has only a single parameter, π_t, their posterior probability that $x_t = 1$. The conditional variance and third moment of x_t, which drive price dynamics, are simple functions of π_t:

$$\kappa_{2,t} = \pi_t (1 - \pi_t) \tag{20}$$

$$\kappa_{3,t} = (1 - 2\pi_t) \times \kappa_{2,t} \tag{21}$$

The variance here then is a bell-shaped function of π_t, peaking at 1/4 at $\pi_t = 1/2$ and declining to zero on both sides, and $sign(\kappa_{3,t}) = sign\left(\frac{1}{2} - \pi_t\right)$. Economically, when π_t is near 1 so that agents are confident they are in the good state, volatility is low, but the third moment is strongly negative, so there is a leverage effect. However, when a bad state is realized and investors have seen enough signals to be confident in that, so that π_t is below 1/2, the leverage effect <u>reverses</u>: agents no longer worry as much about the economy getting worse, so there is relatively more upside and $\kappa_{3,t} > 0$.

These results illustrate the importance, in the context of the leverage effect, of agents continuing to learn in bad states. If learning stops once agents know the economy is in a recession, then the leverage effect disappears or even reverses.

5 Extensions and robustness

5.1 Inferring global properties of beliefs from local information

Proposition 1 says that the local properties of prices are determined by sufficient statistics describing the global properties of beliefs about fundamentals. So far we have analyzed what is required of those global properties in order to generate local price behavior observed empirically. The analysis can also be reversed to show how to use local information about prices to learn about global properties of beliefs.

Under proposition 1, prices follow a diffusion with volatility $\kappa_{2,t}/\sigma_{Y,t}$. There are standard results that then allow for consistent estimation of the diffusive volatility based on high-frequency observations of prices (e.g. Andersen and Teräsvirta (2009)). Those methods therefore allow for real-time estimation of $\kappa_{2,t}/\sigma_{Y,t}$ – a property of beliefs depending on the full probability distribution – without knowing any of the underlying parameters of the model. That is, subject to some potential contamination from $\sigma_{Y,t}$, it is possible to estimate a global feature of beliefs – the variance – in real time using only local variation in prices.

A similar argument holds for the $\kappa_{3,t}$ process. Since the local volatility of prices is $\kappa_{2,t}/\sigma_{Y,t}$, the volatility of volatility, from proposition 1, is $\kappa_{3,t}/\sigma_{Y,t}^2$. There are also nonparametric methods for estimating volatility-of-volatility from high-frequency price data. As with volatility itself, an estimate of volatility-of-volatility – or, for that matter, of the strength of the leverage effect – yields global information about investor beliefs, in this case $\kappa_{3,t}/\sigma_{Y,t}^2$.

5.2 Discrete information revelation events

Dytso, Poor, and Shamai (2022) prove the following discrete version of theorem 1. Instead of assuming a diffusive information flow, this result is for a signal with strictly positive information content (i.e. a positive precision or finite variance), meaning that the cumulants also update by discrete amounts.

Proposition 4 *[Dytso, Poor, and Shamai (2022), equation (52)] For a random variable x_t and a signal $y_t \sim N(x_t, \sigma^2)$,*

$$\frac{d}{dy}\kappa_j\left(x_t \mid y_t = a\right) = \kappa_{j+1}\left(x_t \mid y_t = a\right)/\sigma^2 \tag{22}$$

where $\kappa_j\left(x_t \mid y_t = a\right)$ is the jth posterior cumulant of x_t conditional on observing $y_t = a$.

Furthermore, for y_t in a neighborhood of a,

$$E\left(x_t \mid y_t\right)=\sum_{j=0}^{\infty} \frac{\kappa_{j+1}\left(x_t \mid y_t=a\right)}{j!}\left(\frac{y_t-a}{\sigma^2}\right)^j \tag{23}$$

and an analogous series holds for all higher cumulants.

Proposition 4 shows that the type of recursion in theorem 1 continues to hold for discrete revelation events – diffusive information coming in infinitesimal increments in continuous time is not necessary for the central results. At the same time, it shows that normality is important – we can drop continuity, but proposition 4 still requires a normally distributed signal.[22] That said, proposition 4 also shows why continuous time is useful here: the prior cumulants determine sensitivities, rather than the posterior cumulants that appear in (22). In continuous time the cumulants follow continuous processes, so the prior and posterior values are effectively identical.

5.3 Is it possible to track just a subset of the cumulants?

When looking at theorem 1, a natural question is whether it is possible to ignore the higher-order cumulants and just focus on, say, the first three. The short answer is no. There is no distribution for which there exists an $\bar{n}$ such that $\kappa_n = 0$ for all $n \geq \bar{n}$, except for the normal for which $\bar{n} = 3$. So while it is natural and intuitive, the normal distribution is also an extremely special case, in that there is no other distribution that is even qualitatively similar in terms of the behavior of its higher cumulants.

Additionally, since the gain of the nth cumulant is proportional to the value of the $n+1$th, it cannot be the case that, for example, all the odd cumulants are always equal to zero – the odd cumulants update depending on the values of the even ones, so if any even cumulants are nonzero, then the odd ones are eventually also. Moreover, if any of the higher cumulants is ever nonzero, then the distribution is permanently non-normal (since a Gaussian update of a non-normal distribution always yields a non-normal posterior), and all of the higher cumulants vary over time according to the dynamics in theorem 1.

[22] Dytso and Cardone (2021) explore related results for non-Gaussian variables, but do not derive a power series result. It is possible to derive a similar result for certain other specific cases, e.g. when the likelihood is exponential or Poisson.

6 Illustrative calibration

This section presents a simple quantitative example. We first use it to illustrate the model's core mechanisms with impulse response functions, and then examine the extent to which the qualitative predictions above map into quantitatively reasonable behavior. The example shows how layering incomplete information over simple dynamics for fundamentals can generate severe nonlinearities that help fit a range of features of empirical data on aggregate stock returns. That said, it is important to emphasize that the simulation results are just an example. Their failure to match the data on some dimension does not mean that there is not a model with the sort of learning we have studied so far that would do better, just that the exact specification detailed in this section is (obviously) imperfect.

6.1 Model setup and solution

Fundamentals have an average growth rate of g with occasional exponential downward jumps

$$dx_t = (\phi\lambda + g)\, dt - J_t dN_t \tag{24}$$

where N is a Poisson process with constant rate ϕ and J_t is an exponential random variable with mean λ. $\phi\lambda dt$ ensures that mean price growth is equal to g.[23] The only free parameters determining dynamics are ϕ, λ, and σ_Y; g plays no role except to generate positive average returns.

If the jumps are large and rare, then this can be thought of as a disaster model. Smaller and more frequent jumps might be thought of as representing recessions.

The analytic results in principle require tracking an infinite number of cumulants over time. To simulate, we simply discretize the state space – i.e. treat x_t as discrete Markov chain – and then directly calculate the updates using exact formulas via Bayes' theorem. In the discretization, the fundamental unit of time is taken to be 1/3 of a day and the step size for x is one log point.[24]

[23] The reduced-form process for x can easily be generated by assuming that cash-flows follow the same jump process. Positive risk premia can be generated by assuming the SDF is also driven by the same jumps (but with the opposite exposure).

[24] Note that even if x is discrete, the posterior moments (including prices) are continuous, since they place probabilities (in a continuous interval) on the discrete states. The numerical results are not sensitive to the discretization choices used here.

6.2 Parameter selection

We obtain parameters through simple moment matching – this is not a full-blown estimation exercise. The moments used for fitting are discussed in appendix A.1.5. The parameter values are $\{\phi, \lambda, \sigma_Y\} = \{0.00048, 0.40, 5.04\}$, where the time unit t is taken to measure days. The value of ϕ implies that disasters occur on average once every 8.3 years, and λ implies that the average decline in fundamentals is 40 log points (33 percent). The noise in signals, σ_y, is 504 log points per day. To get a sense of scale, if agents hypothetically had a prior variance for fundamentals of ∞ and fundamentals were constant, after one year of observing such signals their posterior standard deviation would be 0.32 in logs. We set $g = 0.07$ to match the historical equity premium.

6.3 Impulse response functions

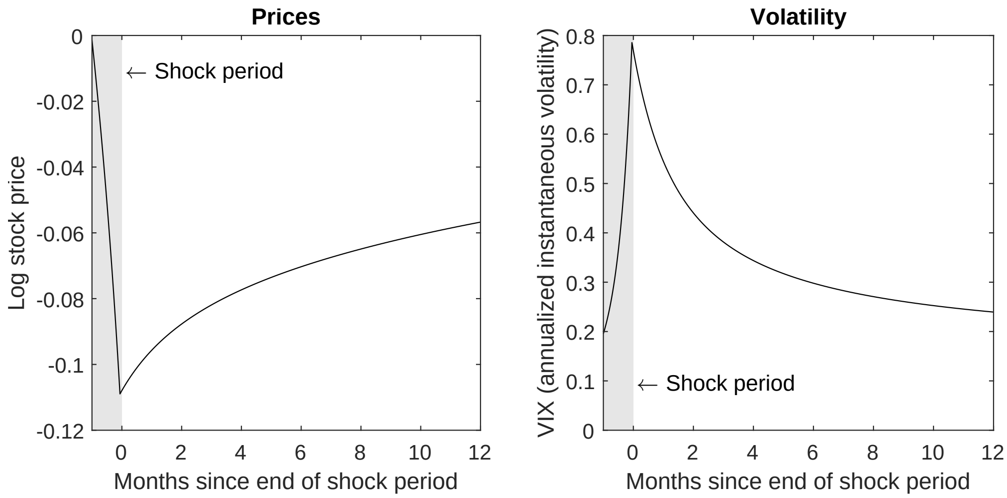
This section examines two impulse response functions – to errors in the signal, $\sigma_Y dW_t$, and to fundamentals, $J_t dN_t$. Since the model is nonlinear, impulse responses differ depending on the state of the economy. We construct the IRFs by first simulating a long period in which agents observe $dY_t = 0$ (i.e. assuming $x_t = 0$ and that agents have just observed the mean signal), and then shocking either the noise or the fundamental.

We calibrate the noise shock to accumulate over the course of one month and represent a 1-in-10 year (2.4-standard deviation) negative surprise. Specifically, over the 21 trading days of the month, agents observe in each period (which is discretized in the simulation) $\Delta Y_t = -2.4\sigma_Y/\sqrt{20}$. Since the model is nonlinear, the size of the shock affects the dynamics, and we choose a large shock to help illustrate the nonlinear effects.

Figure 2 plots the response of prices and volatility to the shock. The first month in the figure is the period in which the shock occurs. As agents observe the negative signals, prices fall and volatility rises. It is possible to see that those effects are convex. Given the results above, as uncertainty (and hence volatility) rises, prices become more sensitive to signals, with the result that the response of prices over the course of the month is concave, with prices declining progressively faster. When the shock ends, prices and volatility revert, but the recovery is very slow: volatility only gets close to its starting point after about a year, and in the same time prices have recovered only about half of their decline. The fact that prices recover initially quickly and then slowly is again a consequence of the volatility dynamics: the recovery of volatility and uncertainty itself slows the recovery in prices.

The second IRF is more standard – just a one-time unit-standard deviation (40 log points) decline in x_t, with $\Delta Y_t = x_t \Delta t$ in each period (i.e. agents see a signal that happens to have

Figure 2: Response to negative error in the signal



Note: The left-hand panel plots the IRF for prices – the conditional expectation of x – and the right hand for the conditional standard deviation of prices. The shock is a one-time unit standard deviation negative error in the signal (i.e. a negative realization of $\sigma_Y dW$.

no noise, but, again, they still update beliefs as though there is noise).

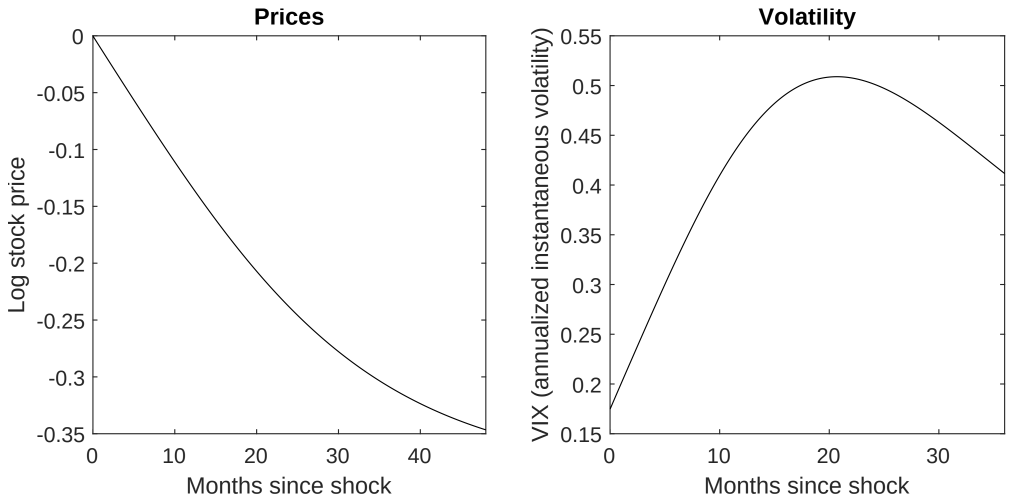
The left-hand plot in figure 3 shows that the decline in prices is again nonlinear – it accelerates before slowing, with the initial declines equal to 0.063 percent per day, accelerating to 0.1 percent per day at their peak. It takes about two years for prices to fully incorporate the drop in fundamentals. Similarly, the peak in volatility actually comes over a year after the occurrence of the shock, with the recovery taking years more. This is therefore a model in which disasters take years on average to fully play out. They are not one-time events, but rather involve rich dynamics, with price declines accelerating and volatility accumulating over time.

6.4 Simulation results

Figure 4 plots the time-series of fundamentals, x, and prices, p, from the full simulation, with the mean growth rate removed to help readability. Prices track fundamentals reasonably well, but clearly there can be large deviations. In some cases, fundamentals jump down and it takes time for prices to catch up, and there are also clear cases of prices jumping down "erroneously" (based on hindsight or on knowing the true state) and then recovering.

Table 1 reports unconditional moments for returns and their conditional volatility. The model broadly matches the data, missing on only the most extreme statistics. Simulated returns have similar volatility and somewhat less skewness than in the data (examined in

Figure 3: Response to a negative realization of fundamentals



Note: These plots are the same as in the previous figure, except they correspond to the IRF for a negative realization of fundamentals. Specifically, the IRF for prices is the average path of prices when $x = -\lambda$ compared to $x = 0$, and the right-hand panel is the same for price volatility.

more detail below), but kurtosis is too small by half. Note that kurtosis is necessarily the most weakly estimated of the three and most strongly driven by outliers. For volatility, the model matches the data very well in levels, and in first differences only misses skewness, which in the data is again driven by a few outliers.

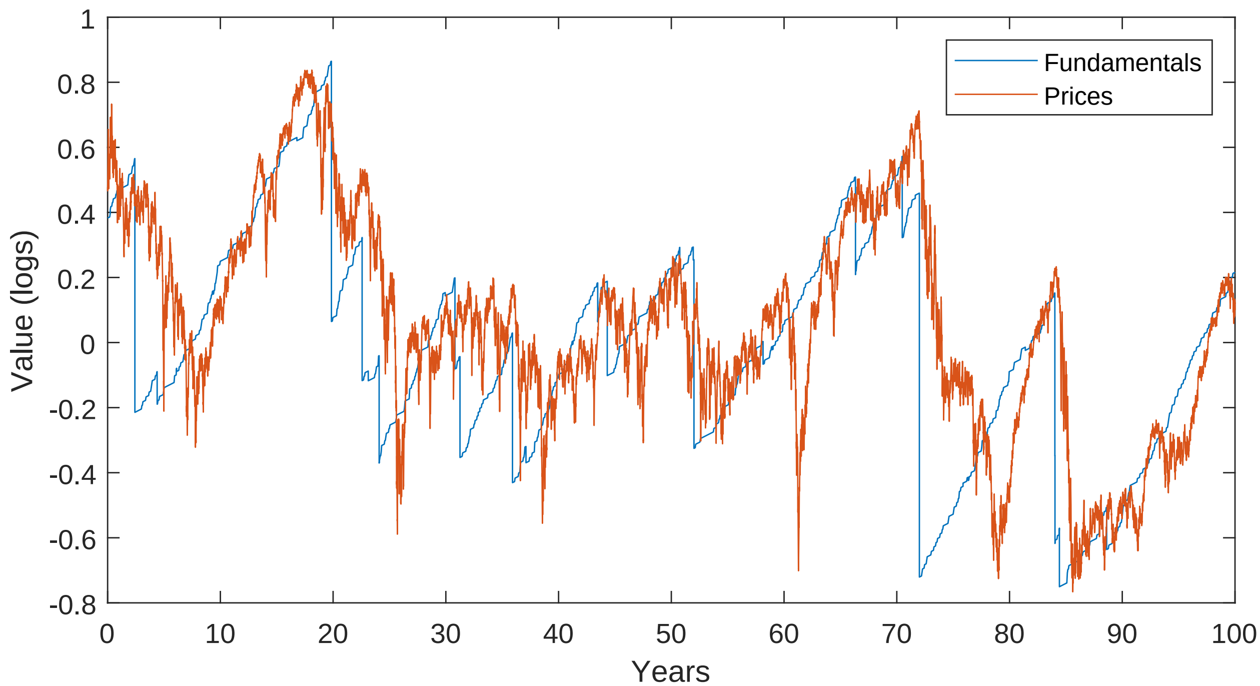
Table 2: Model vs data moments

	Stock returns		Volatility level		Volatility change	
Moment	Data	Model	Data	Model	Data	Model
Std. dev.	1.14	1.19	6.64	8.94	1.39	1.30
Skewness	-0.26	-0.06	2.19	2.17	1.51	-0.06
Kurtosis	12.68	8.13	11.57	10.51	30.03	26.3
Corr. w/ R_t					-0.78	-0.89

Note: This replicates table 1 for the empirical moments and compares them to the model simulation.

Figure 5 compares the model's behavior to what was shown in figure 1. Return skewness is very similar to the data, especially at horizons of a month or more. The inverse autocorrelations of conditional volatility in the model are also essentially identical to the data. The bottom-left panel shows that the model generates a leverage effect scatter plot similar to the data, though with too little dispersion when returns are near zero (implying that there is a component of conditional volatility that is driven by a process that is independent of returns). Finally, the bottom-right panel shows histograms for volatility in the model and

Figure 4: Simulated time series of fundamentals and prices



Note: "Fundamentals" is the simulated x process, and "prices" is the simulated κ_1 process. The mean growth rate has been removed to help make the figure readable.

data demonstrating that the distributions are similar.

The results in this section show that the model is able to match key features of the data not just qualitatively but also quantitatively.

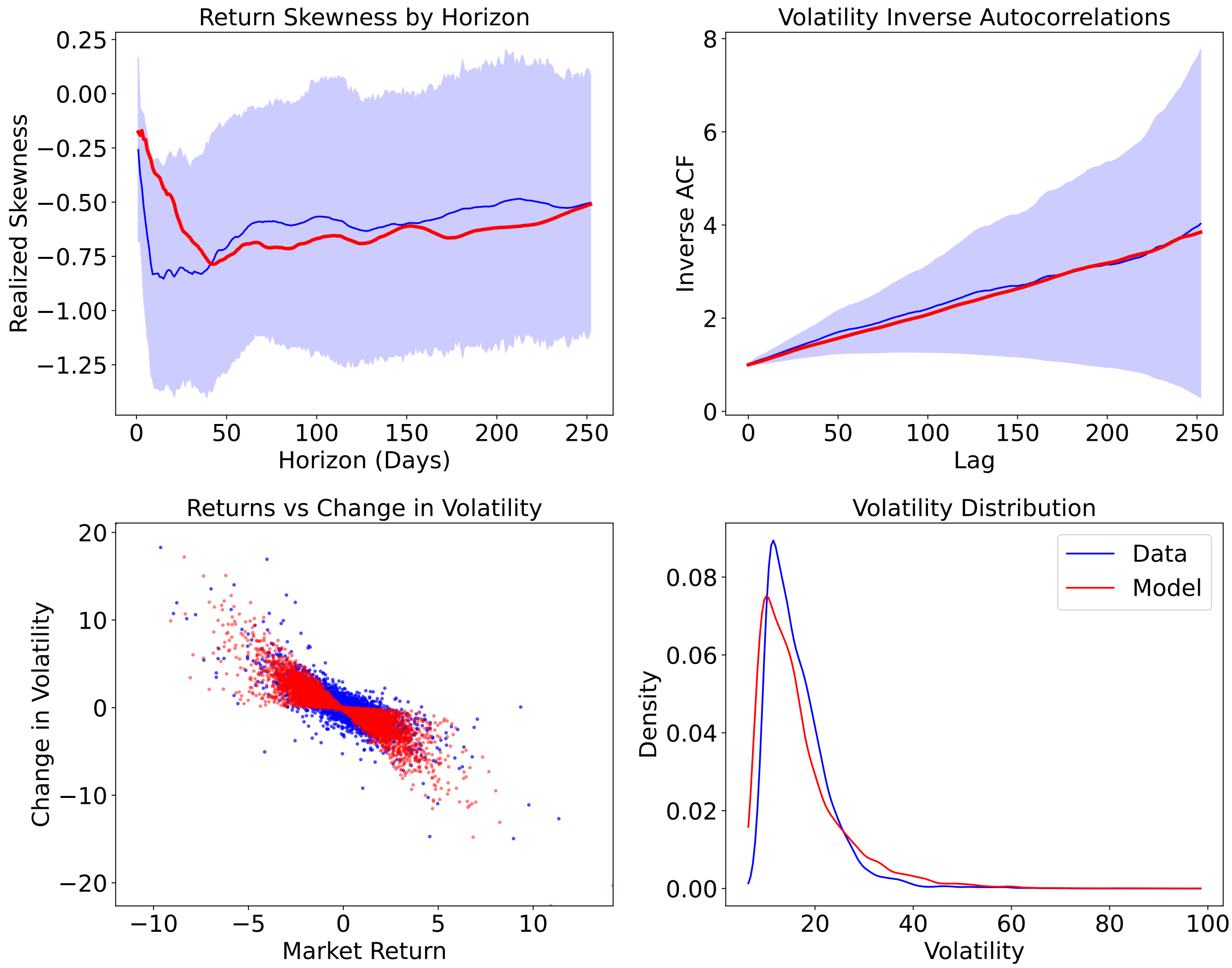
7 Estimated volatility dynamics and investor uncertainty

The analytic results in section 4 have specific implications for the dynamics of volatility and the leverage effect. This section focuses on estimating the regressions motivated by the model-implied dynamics for volatility. They deliver two key outputs: estimates of the noise in investors' signals and tests of overidentifying restrictions.

7.1 Regression setup

Combining equations (13), (8), and (17) (and ignoring higher-order error in skewness and holding $\sigma_{Y,t}$ constant), we have

Figure 5: Simulation results



Note: The top left, top right and bottom left panels are the same as in figure 1, overlayed with the corresponding outputs from a 100-year simulation of the model based on the parameters discussed in the text. The bottom-right panel shows the estimated density of volatility in the data and in the model.

$$\begin{aligned} d\left(vol_t\right) &= \left(\frac{1}{3}\Delta t^{-1/2}\right) skew_t\left(p_{t+\Delta t}\right) dp_t - \frac{1}{\sigma_{Y,t}} vol_t^2 dt \\ &\quad + E_t\left[d\left[x\right]_t\right] + 2\operatorname{cov}_t\left(x_t, dx_t\right) \end{aligned} \tag{25}$$

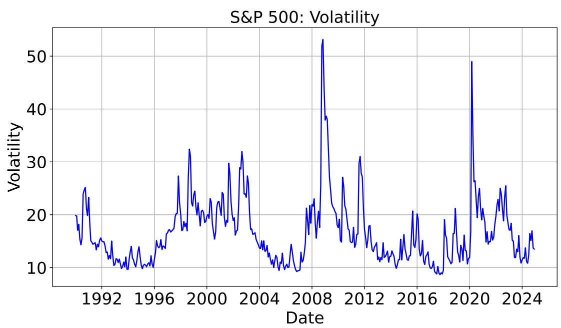
If x has independent increments – as in the quantitative model – then the second line reduces to a constant. We take that as our benchmark in this section.

The model has two predictions for the results of this regression: the coefficient on $\left(\frac{1}{3}\Delta t^{-1/2}\right) skew\left(p_{t+\Delta t}\right) dp_t$ should be equal to 1, and the coefficient on $vol_t^2 dt$ is equal to $\sigma_{Y,t}^{-1}$. The first represents a test of the model. The second shows that the regression can be used to identify one of the model's structural parameters.

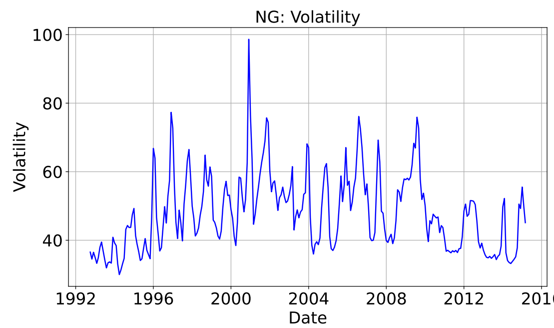
7.2 Data

We estimate the regression (25) for two markets. The first, given our focus on the stock market, is the S&P 500. For that case, we proxy for $std\left(dp_t\right)$ with the same conditional volatility measure that we have used throughout. For the return dp_t we continue to use the log return on the CRSP total market index. Last, similar to volatility, we construct $skew_t\left(p_{t+\Delta t}\right)$ by directly forecasting the realized second and third moments. The top-left panel of figure 6 plots the measure of conditional volatility and the bottom-left conditional skewness over our sample period.

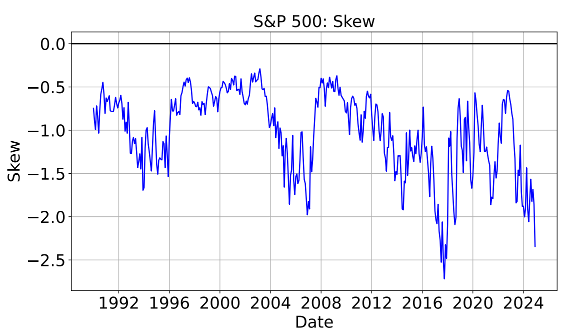
Figure 6: Time series of VIX and Skew for S&P 500 and natural gas



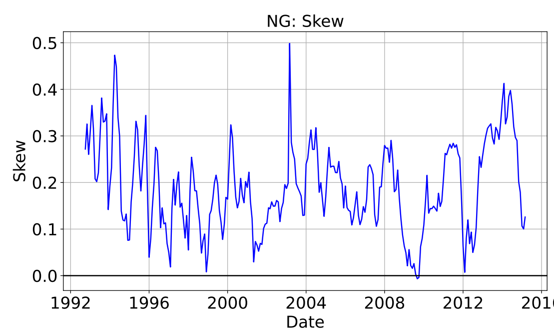
(a) Volatility (S&P 500)



(b) Volatility (natural gas)



(c) Skew (S&P 500)



(d) Skew (natural gas)

Note: Time series plots for the volatility and skewness for the S&P 500 and for natural gas. All series are obtained from regressing realized moments onto the corresponding option-implied moments, and using fitted values to account for risk premia in options. All series are monthly average of the corresponding daily series.

The S&P 500 conditional skewness is almost exclusively negative, so for the second market to use for estimation, we choose natural gas because it is a large and economically significant market that displays consistently positive skewness. To calculate dp_t in this case we use natural gas futures returns from the CME. We then use options on futures to estimate the conditional moments in the same was as for the S&P 500.[25] The time series of conditional volatility and skewness for natural gas are plotted in the right-hand panels of figure 6. Due to the seasonality in natural gas prices, we include contract fixed effects in the volatility regressions for natural gas.

In addition to the differences in skewness, the S&P 500 is generally viewed as being significantly more important systematically – since it represents a nontrivial part of aggregate wealth – so if one was concerned about the results for the S&P 500 being somehow contaminated by risk premia, that might be less of a concern for natural gas.

There are many other underlyings that could be studied, like individual stocks, bonds, and other commodities. The two markets here are simply meant to illustrate the model's core mechanisms and show that, as a first pass, they map somewhat reasonably into the data.

7.3 Results

Table 3 reports results of the regression implied by (25) for the S&P 500 and natural gas. The first and third columns report the baseline results The coefficients are highly statistically significant and have the expected sign.

Under the model, if the various assumptions we made to derive the regression here are true, the coefficient on $\frac{1}{3\sqrt{21}} skew_t dp_t$ should be equal to 1, and in both cases that is a very good description of the data. The coefficient is 0.97 with a (robust) standard error of 0.03 for the S&P 500 and 0.97 with a standard error of 0.08 for natural gas. There are really two key features of the model that generate that prediction: that returns and their volatility are jointly driven by the same shock (i.e. the same Brownian motion), and that they jointly follow a diffusion.

To evaluate the relevance of the model-implied nonlinear in volatility dynamics, the second and fourth columns of table 3 include both vol_{t-1}^2 and vol_{t-1}. vol_{t-1}^2 does in fact appear to dominate. In both cases, the t-statistic for vol_{t-1}^2 is larger than that on vol_{t-1} – for the S&P by a factor of more than 10, but only marginally so for natural gas – indicating that vol_{t-1}^2 has greater explanatory power than vol_{t-1}.

The coefficients on vol_{t-1}^2 give an estimate of σ_Y^{-1} at the daily level (given that these are daily regressions). We further evaluate the implication of those estimates below.

[25] The specific methods are from Dew-Becker (2024).

Table 3: Volatility regressions

	S&P 500			Natural Gas		
	(1)	(2)	(3)	(4)	(5)	(6)
	S&P 500			Natural Gas		
	(1)	(2)	(3)	(4)	(5)	(6)
Vol^2_{t-1}	-0.74***	-0.79	-0.58***	-0.20***	-1.00*	-0.20***
	[0.16]	[0.52]	[0.14]	[0.07]	[0.56]	[0.07]
$\frac{1}{3\sqrt{21}} skew_{t-1} dp_t$	0.97***	0.97***	0.24***	0.97***	0.96***	0.64***
	[0.03]	[0.03]	[0.04]	[0.08]	[0.08]	[0.15]
Vol_{t-1}		0.001			0.056	
		[0.013]			[0.036]	
dp_t			-0.048***			0.005***
			[0.003]			[0.002]
R^2	0.54	0.54	0.63	0.12	0.12	0.12

Note: Daily regressions of first differences in volatility (for S&P 500 and natural gas) onto different predictors. Stars indicate statistical significance: * p<.1, ** p<.05, ***p<.01.

7.4 Estimates of investors' uncertainty about fundamentals

Having estimates for σ_Y allows us to then use the volatility and skewness of stock market returns to reveal the standard deviation and skewness of agents' posteriors for fundamentals. Specifically, recall that

$$vol_t = \frac{\kappa_{2,t}}{\sigma_{Y,t}} \Delta t^{1/2} \tag{26}$$

$$\Rightarrow \kappa_{2,t}^{1/2} = \left(vol_t \sigma_{Y,t} \Delta t^{-1/2}\right)^{1/2} \tag{27}$$

Recall that the scaling of the estimates is for a unit time interval being equal to a day. The daily standard deviation of stock returns in our sample is 1.05 percent. If $\sigma_{Y,t}$ is between 0.94 and 2.36, based on the coefficient on vol^2_{t-1} in equation (25), that implies that agents' posterior standard deviation is between 10.4 and 16.5 percent. The ± 2 standard deviation range for fundamentals around the current price for the aggregate stock market is then between ± 20.8 and ± 33.0 percent.

Similarly, we can get an estimate of average skewness in beliefs. One-month conditional return skewness is historically approximately -1. Plugging that into (17) along with the estimates of κ_2 and σ_Y yields an estimate for the skewness of fundamentals between -0.29 and -1.13. In the time series, the estimate of conditional skewness of fundamentals is proportional to the conditional skewness of returns divided by the square root of the conditional standard deviation of returns.

We have not found a survey that directly measures investors' uncertainty about fundamentals (e.g. that asks them about probabilities that the fundamental value might fall in different ranges, as the Survey of Consumer Expectations and Survey of Professional Forecasters do for inflation and other variables). However, uncertainty is sometimes proxied for by disagreement – and it is plausible that they are at least somewhat related – so a survey giving a cross-section of estimates of fundamental value would be one way to sanity check for the model-implied estimate of average uncertainty.

The Investor Behavior Project at Yale has a survey of institutional investors that asks the following question: "What do you think would be a sensible level for the Dow Jones Industrial Average based on your assessment of U.S. corporate strength (fundamentals)?" We interpret the answer to that question as each investor's estimate of $E\left[\exp\left(x_t\right) \mid Y^t\right]$. To calculate cross-sectional dispersion, given that the surveys are completed on different dates by different respondents, we calculate the average squared log difference between each investor's reported fundamental value and the actual value at the time of the survey. The square root of that average represents a measure of the cross-sectional standard deviation.

The data runs from August, 1993 to July, 2024 and has 8,242 observations. In that sample, we estimate the cross-sectional standard deviation to be 17.0 percent, which lies just outside the edge of the confidence bands for $\kappa_2^{1/2}$. Again, while disagreement and uncertainty are different concepts, it is plausible that degree of disagreement across people would be of a similar magnitude to overall uncertainty, and we observe that here.

7.5 Summary

Overall, this section shows that the model's predictions for volatility dynamics match the data well, both for the S&P 500 and natural gas futures. The prediction for nonlinear mean reversion – via a quadratic term in the regression – is well confirmed, and in fact it drives out a linear mean reversion term. The prediction that market returns should be interacted with a measure of skewness appears not inconsistent with the data, but it is also not dominant – raw returns themselves are still a significant predictor of the change in conditional volatility.

Finally, the coefficients themselves can be mapped into an estimate of σ_Y, the noise in investors' signals. The model implies that the rate of mean reversion depends on that noise, and the estimated confidence interval for that quantity for the S&P 500, [1.26, 3.10], accords well with the value that we also find works well in the calibration. That estimate then also implies that investors' uncertainty about the true fundamental value of stocks – if they had complete information – is $\pm 22 - 35$ percent. Moreover, the implied uncertainty matches well with the Yale IBP survey measure of cross-sectional disagreement.

8 Conclusion

This paper's main results are fundamentally about how information affects the various moments of agents' beliefs in a very simple but standard Bayesian filtering setting. The analysis is motivated by behavior of the stock market, and the analysis shows both that the theoretical results can help elucidate one mechanism that generates comovements among many higher moments of returns, and also that the mechanism can generate quantitatively reasonable behavior.

But the general model setup that we solve is certainly not applicable just to the aggregate stock market. The results have implications for beliefs in any setting, whether that be other financial markets, surveys, or competitive settings.

References

Abel, Andrew B., Janice C. Eberly, and Stavros Panageas, "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," *Econometrica*, 2013, *81* (4), 1455–1481.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Brent Meyer, and Nicholas Parker, "Surveying business uncertainty," *Journal of Econometrics*, 2022, *231* (1), 282–303.

Andersen, Torben and Timo Teräsvirta, *Realized Volatility* 2009.

Bachmann, Rdiger, Kai Carstensen, Stefan Lautenbacherr, Manuel Menkhoff, and Martin Schneider, "Uncertainty and Change: Survey Evidence of Firms Subjective Beliefs," 2024. Working paper.

Bain, Alan and Dan Crisan, *Fundamentals of stochastic filtering*, Vol. 3 2009.

Baker, Scott, Nicholas Bloom, Steven J. Davis, and Marco Sammon, "What Triggers Stock Market Jumps?," 2025. Working paper.

Barro, Robert J., "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 2006, *121(3)*, 823–866.

Berger, David, Ian Dew-Becker, and Stefano Giglio, "Uncertainty shocks as second-moment news shocks," *The Review of Economic Studies*, 2020, *87* (1), 40–76.

Bianchi, Francesco, Cosmin Ilut, and Martin Schneider, "Uncertainty shocks, asset supply and pricing over the business cycle," 2017. Working paper.

Bianchi, Javier, "Overborrowing and systemic externalities in the business cycle," *American Economic Review*, 2011, *101* (7), 3400–3426.

Bollerslev, Tim and Hans Ole Mikkelsen, "Modeling and pricing long memory in stock market volatility," *Journal of econometrics*, 1996, *73* (1), 151–184.

_ , George Tauchen, and Hao Zhou, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 2009, *22(11)*, 4463–4492.

Breeden, Douglas T and Robert H Litzenberger, "Prices of state-contingent claims implicit in option prices," *Journal of business*, 1978, pp. 621–651.

Britten-Jones, Mark and Anthony Neuberger, "Option prices, implied price processes, and stochastic volatility," *The journal of Finance*, 2000, *55* (2), 839–866.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer, "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *The American Economic Review*, 2016, *106* (3), 664–698.

Comtet, Louis, *Advanced Combinatorics: The art of finite and infinite expansions*, Springer Science & Business Media, 1974.

Corsi, Fulvio, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 2009, *7* (2), 174–196.

Cox, John C., Jr. Jonathan E. Ingersoll, and Stephen A. Ross, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 1985, *53(2)*, 385–407.

David, Alexander, "Fluctuating confidence in stock markets: Implications for returns and volatility," *Journal of Financial and Quantitative Analysis*, 1997, *32* (4), 427–462.

_ and Pietro Veronesi, "What ties return volatilities to price valuations and fundamentals?," *Journal of Political Economy*, 2013, *121* (4), 682–746.

Dew-Becker, Ian, "Real-time forward-looking skewness over the business cycle," 2024.

Ding, Zhuanxin, Clive WJ Granger, and Robert F Engle, "A long memory property of stock market returns and a new model," *Journal of empirical finance*, 1993, *1* (1), 83–106.

Dupraz, Stéphane, Emi Nakamura, and Jón Steinsson, "A plucking model of business cycles." Working paper.

Dytso, Alex and Martina Cardone, "A general derivative identity for the conditional expectation with focus on the exponential family," in "2021 IEEE Information Theory Workshop (ITW)" IEEE 2021, pp. 1–6.

_ , H Vincent Poor, and Shlomo Shamai Shitz, "Conditional mean estimation in Gaussian noise: A meta derivative identity with applications," *IEEE Transactions on Information Theory*, 2022, *69* (3), 1883–1898.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson, "Learning about the long run," *Journal of Political Economy*, 2024, *132* (10), 000–000.

French, Kenneth R, G William Schwert, and Robert F Stambaugh, "Expected stock returns and volatility," *Journal of financial Economics*, 1987, *19* (1), 3–29.

Gabaix, Xavier, "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 2012, *127(2)*, 645–700.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Neglected risks: The psychology of financial crises," *American Economic Review*, 2015, *105* (5), 310–314.

Gilchrist, Simon and John C Williams, "Putty-clay and investment: a business cycle analysis," *Journal of Political Economy*, 2000, *108* (5), 928–960.

Granger, Clive WJ, "Long memory relationships and the aggregation of dynamic models," *Journal of econometrics*, 1980, *14* (2), 227–238.

Guo, Dongning, Yihong Wu, Shlomo Shamai, and Sergio Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, 2011, *57* (4), 2371–2385.

Hansen, Gary D and Edward C Prescott, "Capacity constraints, asymmetries, and the business cycle," *Review of Economic Dynamics*, 2005, *8* (4), 850–865.

Ilut, Cosmin, Matthias Kehrig, and Martin Schneider, "Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News," *Journal of Political Economy*, 2018, *126* (5), 2011–2071.

Johannes, Michael, Lars A Lochstoer, and Yiqun Mou, "Learning about consumption dynamics," *The Journal of finance*, 2016, *71* (2), 551–600.

Kocherlakota, Narayana, "Creating business cycles through credit constraints," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2000, *24* (3), 2–10.

Kozeniauskas, Nicholas, Anna Orlik, and Laura Veldkamp, "What are uncertainty shocks?," *Journal of Monetary Economics*, 2018, *100*, 1–15.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran, "The tail that keeps the riskless rate low," *NBER Macroeconomics Annual*, 2018, *33* (1), 253–283.

Lettau, Martin and Sydney Ludvigson, "Measuring and Modeling Variation in the Risk-Return Tradeoff," in Yacine Ait-Sahalia and Lars P. Hansen, eds., *Handbook of Financial Econometrics*, Vol. 1, Elsevier Science B.V., North Holland, Amsterdam, 2010, pp. 617–690.

Liptser, Robert S and Albert N Shiryaev, *Statistics of Random Processes: I. General Theory*, Springer Science & Business Media, 2013.

Lukacs, Eugene, *Characteristic Functions*, second ed., New York: Hafner Publishing Company, 1970.

Mandelbrot, Benoit, "The Variation of Certain Speculative Prices," *Journal of Business*, 1963, *36* (4), 394–419.

_ **, Adlai Fisher, and Laurent Calvet**, "A Multifractal Model of Asset Returns," 1997. Working paper.

McKay, Alisdair and Ricardo Reis, "The brevity and violence of contractions and expansions," *Journal of Monetary Economics*, 2008, *55* (4), 738–751.

Merton, Robert C, "On estimating the expected return on the market: An exploratory investigation," *Journal of financial economics*, 1980, *8* (4), 323–361.

Moreira, Alan and Tyler Muir, "Volatility-managed portfolios," *The Journal of Finance*, 2017, *72* (4), 1611–1644.

Morley, James and Jeremy Piger, "The asymmetric business cycle," *Review of Economics and Statistics*, 2012, *94* (1), 208–221.

Neuberger, Anthony, "Realized skewness," *The Review of Financial Studies*, 2012, *25* (11), 3423–3455.

_ **and Richard Payne**, "The skewness of the stock market over long horizons," *The Review of Financial Studies*, 2021, *34* (3), 1572–1616.

Orlik, Anna and Laura Veldkamp, "Understanding uncertainty shocks and the role of black swans," *Journal of Economic Theory*, 2024, p. 105905.

Rietz, Thomas A., "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 1988, *22(1)*, 117–131.

Sichel, Daniel E, "Business cycle asymmetry: a deeper look," *Economic inquiry*, 1993, *31* (2), 224–236.

Straub, Ludwig and Robert Ulbricht, "Endogenous Second Moments: A Unified Approach to Fluctuations in Risk, Dispersion and Uncertainty," *183*, 625–660.

Veronesi, Pietro, "Stock market overreactions to bad news in good times: a rational expectations equilibrium model," *The Review of Financial Studies*, 1999, *12* (5), 975–1007.

Wachter, Jessica A., "Can time-varying risk of rare disasters explain aggregate stock market volatility?," *Journal of Finance*, 2013, *68(3)*, 987–1035.

Wachter, Jessica and Yicheng Zhu, "Learning with Rare Disasters," 2023. Working paper.

Weitzman, Martin L, "Subjective expectations and asset-return puzzles," *American Economic Review*, 2007, *97* (4), 1102–1130.

A.1 Proofs

A.1.1 Theorem 1

A.1.1.1 Assumptions

Assumption 1 *Let $\phi_{\omega,t} = \exp(i\omega x_t)$ denote the complex exponential of x_t. For any ω, there exists an adapted process $\mathcal{G}_{\omega,s}$ satisfying, a.s., $\int_0^t |\mathcal{G}_{\omega,s}| ds < \infty$ and $\int_0^t \mathbb{E}[\mathcal{G}_{\omega,s}^2] ds < \infty$, such* *that* $\phi_{\omega,t} - \phi_{\omega,0} - \int_0^t \mathcal{G}_{\omega,s} ds$ *is a right-continuous martingale.*

Assumption 2 $\int_0^t \mathbb{E}[x_s^2]ds < \infty$ *and* $\int_0^t |x_s|ds < \infty$ *almost surely.*

Assumption 3 *The process* $\sigma_{Y,t}$ *is progressively measurable with respect to the natural filtration of* Y_t*. Furthermore,*

$$\mathbb{P}\left(\int_0^t \sigma_{Y,s}^2 ds < \infty\right) = 1, \tag{A.1}$$

$$0 < \underline{\sigma}^2 \leq \sigma_{Y,t}^2, \tag{A.2}$$

$$\left|\sigma_{Y,t} - \sigma_{\tilde{Y},t}\right|^2 \leq L_1 \int_0^t (Y_s - \tilde{Y}_s)^2 dK(s) + L_2(Y_t - \tilde{Y}_t)^2, \tag{A.3}$$

$$\sigma_{Y,t}^2 \leq L_1 \int_0^t (1 + Y_s^2) dK(s) + L_2(1 + Y_t^2), \tag{A.4}$$

where L_1 *and* L_2 *are non-negative constants and* $K(t)$ *is a non-decreasing right-continuous function satisfying* $0 \leq K(t) \leq 1$ *for all* $t < \infty$*.*

A.1.1.2 Proof

Lemma 1 *Let* $\varphi_{x,t}(\omega) = \mathbb{E}[\exp(i\omega x_t)|Y^t]$ *denote the characteristic function of the posterior distribution of* x_t *conditional on* Y^t*. If assumptions 1–3 are satisfied, then*

$$d\varphi_{x,t}(\omega) = \mathbb{E}_t[d\exp(i\omega x_t)] + \mathrm{cov}_t(x_t, \exp(i\omega x_t)) \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_{Y,t}^2}, \tag{A.5}$$

where $\mathbb{E}_t$ *and* cov_t *denote the expectation and covariance operators, respectively, conditional on* Y^t*.*

The lemma follows from theorem 8.1 of Liptser and Shiryaev (2013) by setting $h_t \to \phi_{\omega,t}$, $\xi_t \to Y_t$, $A_t \to x_t$, and $B_t(\xi) \to \sigma_{Y,t}$.[1] We proceed by verifying that conditions (8.1)–(8.9) of Liptser and Shiryaev (2013) are satisfied.

Equation (8.2) is simply equation (6) of the paper in integral form. Assumption 1 implies that conditions (8.1) and (8.7) are satisfied. The first part of condition (8.3) and condition (8.8) are satisfied by assumption 2. The second part of condition (8.3) and conditions (8.4), (8.5), (8.9) are satisfied by assumption 3. Condition (8.6) is satisfied since $\phi_{\omega,t}$ is a

[1]The result is stated for real-valued functions. However, it can trivially be extended to the complex-valued function $x \mapsto \exp(i\omega x)$ using the identity $\exp(i\omega x) = \cos(\omega x) + i\sin(\omega x)$ and separately considering the real and imaginary parts of the function.

bounded function. Finally, applying theorem 8.1 and noting that the Brownian motion W_t is independent of x_t, we get

$$\mathbb{E}_t[\exp(i\omega x_t)] = \mathbb{E}_0[\exp(i\omega x_0)] + \int_0^t \mathbb{E}_s[\mathcal{G}_{\omega,s}]ds + \int_0^t \frac{\mathrm{cov}_s\left(x_s, \exp(i\omega x_s)\right)}{\sigma_{Y,s}} d\overline{W}_s, \tag{A.6}$$

where

$$\overline{W}_t = \int_0^t \frac{dY_s - \mathbb{E}_s[x_s]ds}{\sigma_{Y,s}}. \tag{A.7}$$

Or equivalently,

$$d\mathbb{E}_t[\exp(i\omega x_t)] = \mathbb{E}_t[\mathcal{G}_{\omega,t}]dt + \mathrm{cov}_t(x_t, \exp(i\omega x_t)) \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma^2_{Y,t}}. \tag{A.8}$$

On the other hand, by the definition of $\mathcal{G}_{\omega,t}$,

$$d \exp(i\omega x_t) - \mathcal{G}_{\omega,t}dt = dM_t, \tag{A.9}$$

where M_t is a right-continuous martingale. Therefore, $\mathbb{E}_t[\mathcal{G}_{\omega,t}]dt = \mathbb{E}_t[d \exp(i\omega x_t)]$.

Theorem 4 *Let $\kappa_{k,t}$ denote the kth cumulant of the posterior distribution of x_t conditional on Y^t. Suppose the $n+1$th moment of the posterior distribution and $\mathbb{E}_t[d(x_t^n)]$ both exist, and assumptions 1–3 are satisfied. Then for every $k \leq n$,*

$$\begin{aligned} d\kappa_{k,t} = & \sum_{j=1}^{k} \binom{k}{j} B_{k-j}\left(-\kappa_{1,t}, \ldots, -\kappa_{k-j,t}\right) \mathbb{E}_t[d(x_t^j)] + \frac{\kappa_{k+1,t}}{\sigma^2_{Y,t}} \left(dY_t - \mathbb{E}_t[x_t]dt\right) \\ & - \frac{1}{2\sigma^2_{Y,t}} \sum_{j=2}^{k} \binom{k}{j-1} \kappa_{j,t}\kappa_{k-j+2,t}dt, \end{aligned} \tag{A.10}$$

where B_j denotes the jth complete exponential Bell polynomial.

The result follows from applying Itô's lemma to lemma 1, yielding

$$\begin{aligned} d \log \varphi_{x,t}(\omega) = & \frac{\mathbb{E}_t[d \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} + \frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma^2_{Y,t}} \\ & - \frac{1}{2\sigma^2_{Y,t}} \left(\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} \right)^2 dt \end{aligned} \tag{A.11}$$

and then taking derivatives of both sides.

We begin by verifying existence. By assumption, the posterior distribution of x_t conditional on Y^t has $n+1$ moments. Therefore, the posterior also has $n+1$ cumulants, the corresponding characteristic function has $n+1$ derivatives at $\omega = 0$, and the cumulants are related to the derivatives of the characteristic function through

$$\kappa_{k,t} = i^{-k} \frac{d^k}{d\omega^k} \log \varphi_{x,t}(\omega)\bigg|_{\omega=0} \tag{A.12}$$

for any $k \leq n+1$.[2] Differentiating the left-hand side of (A.11) with respect to ω (and applying the dominated convergence theorem),

$$d\left(\frac{d^k}{d\omega^k} \log \varphi_{x,t}(\omega)\bigg|_{\omega=0} \right) = i^k d\kappa_{k,t}. \tag{A.13}$$

The remainder of the proof calculates the kth derivative of the right-hand side of (A.11) for $k \leq n+1$.

For the **first term**, since x_t has n moments, for any ω in a sufficiently small neighborhood of the origin,

$$\mathbb{E}_t[d\exp(i\omega x_t)] = \sum_{j=0}^{n+1} \frac{(i\omega)^j}{j!} \mathbb{E}_t[d(x_t^j)] + o(\omega^{n+1}). \tag{A.14}$$

Therefore,

$$\frac{d^k}{d\omega^k} \mathbb{E}_t[d\exp(i\omega x_t)]\bigg|_{\omega=0} = i^k \sum_{j=k}^{n+1} \frac{(i\omega)^{j-k}}{(j-k)!} \mathbb{E}_t[d(x_t^j)]\bigg|_{\omega=0} = i^k \mathbb{E}_t[d(x_t^k)]. \tag{A.15}$$

The Leibniz rule then yields

$$\frac{d^k}{d\omega^k} \frac{\mathbb{E}_t[d\exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} = \sum_{j=0}^{k} \binom{k}{j} \left(\frac{d^{k-j}}{d\omega^{k-j}} \mathbb{E}_t[d\exp(i\omega x_t)]\bigg|_{\omega=0} \right) \left(\frac{d^j}{d\omega^j} \left(\mathbb{E}_t[\exp(i\omega x_t)]\right)^{-1}\bigg|_{\omega=0} \right).$$

Finally, note that $\left(\mathbb{E}_t[\exp(i\omega x_t)]\right)^{-1} = \exp\left(-\log \varphi_{x,t}(\omega)\right)$, and the complete exponential Bell polynomials give (see Comtet (1974) section 3.3)

$$\begin{aligned} \frac{d^k}{d\omega^k} \frac{\mathbb{E}_t[d\exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} &= i^k \sum_{j=0}^{k} \binom{k}{j} B_j\left(-\kappa_{1,t}, \ldots, -\kappa_{j,t}\right) \mathbb{E}_t[d(x_t^{k-j})] && \text{(A.16)} \\ &= i^k \sum_{j=1}^{k} \binom{k}{j} B_{k-j}\left(-\kappa_{1,t}, \ldots, -\kappa_{k-j,t}\right) \mathbb{E}_t[d(x_t^{j})]. && \text{(A.17)} \end{aligned}$$

[2] All the results on characteristic functions, moments, and cumulants used here can be found in Chapter 2 of Lukacs (1970).

For the **second term** on the right-hand side of (A.11), we have

$$\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} = \frac{\mathbb{E}_t[x_t \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} - \mathbb{E}_t[x_t] = i^{-1}\frac{d}{d\omega}\log\varphi_{x,t}(\omega) - \mathbb{E}_t[x_t]. \tag{A.18}$$

Therefore,

$$\frac{d^k}{d\omega^k}\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} = i^{-1}\frac{d^{k+1}}{d\omega^{k+1}}\log\varphi_{x,t}(\omega)\bigg|_{\omega=0} = i^k\kappa_{k+1,t}, \tag{A.19}$$

and the kth derivative of the second term in (A.11), evaluated at $\omega = 0$, is given by

$$i^k\frac{\kappa_{k+1,t}}{\sigma^2_{Y,t}}\left(dY_t - \mathbb{E}_t[x_t]dt\right). \tag{A.20}$$

Finally, for the **third term** in (A.11), the Leibniz rule combined with the results above gives

$$\frac{d^k}{d\omega^k}\left(\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\right)^2\Bigg|_{\omega=0} \tag{A.21}$$

$$= \sum_{j=1}^{k-1}\binom{k}{j}\left(\frac{d^j}{d\omega^j}\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0}\right)\left(\frac{d^{k-j}}{d\omega^{k-j}}\frac{\mathrm{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0}\right) \tag{A.22}$$

$$= \sum_{j=1}^{k-1}\binom{k}{j} i^j\kappa_{j+1,t} i^{k-j}\kappa_{k-j+1,t} \tag{A.23}$$

$$= i^k\sum_{j=2}^{k}\binom{k}{j-1}\kappa_{j,t}\kappa_{k-j+2,t}. \tag{A.24}$$

Putting everything together and canceling the i^k constants completes the proof of the theorem.

A.1.2 Derivation of equation (17)

Doing a second order approximation of the price process using proposition 1,

$$\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2] = \kappa^2_{2,t}\sigma^{-2}_{Y,t}\Delta t + O(\Delta t^2), \tag{A.25}$$

and

$$\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3] = \mathbb{E}_t\left[\left(\kappa_{2,t}\sigma_{Y,t}^{-1}\Delta W_t + \frac{\kappa_{3,t}\sigma_{Y,t}^{-1}}{2}(\Delta W_t^2 - \Delta t)\right)^3\right] + O(\Delta t^{5/2}) \tag{A.26}$$

$$= 3\kappa_{2,t}^2\kappa_{3,t}\sigma_{Y,t}^{-4}\Delta t^2 + O(\Delta t^{5/2}). \tag{A.27}$$

Therefore,

$$skew_t\left(p_{t+\Delta t}\right)(\Delta t)^{-1/2} = \frac{\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3]\,(\Delta t)^{-1/2}}{\left(\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2]\right)^{3/2}} \tag{A.28}$$

$$= \frac{\kappa_{3,t}}{\kappa_{2,t}}\sigma_{Y,t}^{-1} + O(\Delta t^{1/2}). \tag{A.29}$$

A.1.3 Derivation of equation (25)

Starting from equation (8),

$$d\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right) = \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.30}$$

$$= \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\kappa_{2,t}}\frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.31}$$

$$= \frac{1}{\sigma_{Y,t}^2}\left(skew_{t\to t+\Delta t}\left(dp_t\right)(\Delta t)^{-1/2}\right)\frac{1}{3}\frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.32}$$

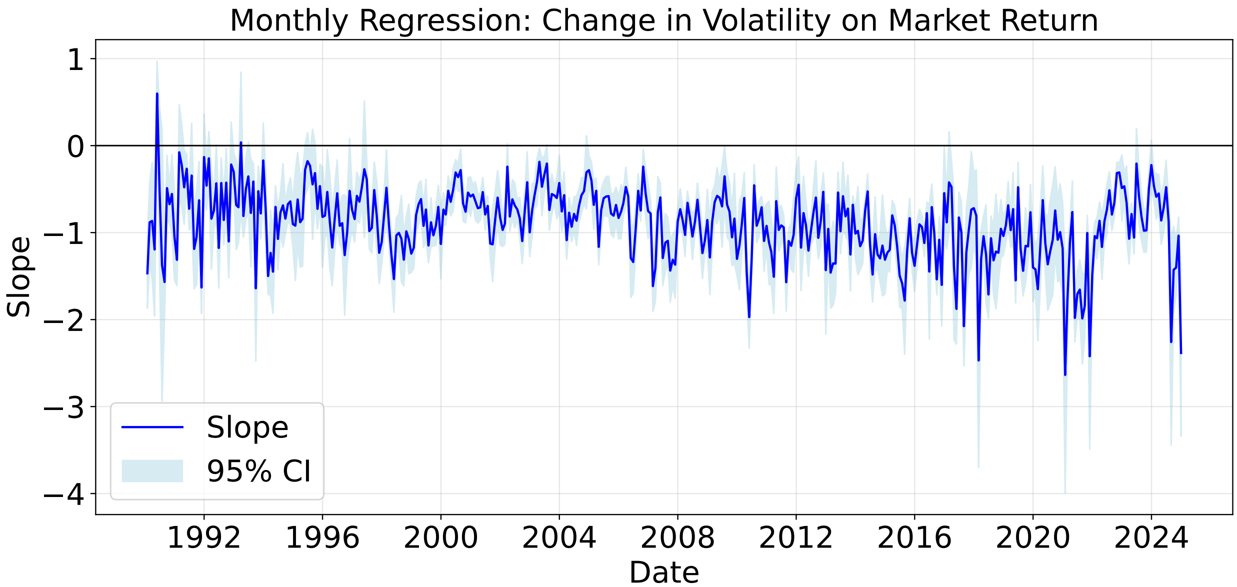
$$= \frac{1}{\sigma_{Y,t}^2}\left(skew_{t\to t+\Delta t}\left(dp_t\right)(\Delta t)^{-1/2}\right)\frac{1}{3}\left[dp_t - E_t dx_t\right] + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt, \tag{A.33}$$

where the third line uses equation (17) and the fourth line inserts the formula for $dp_t = d\kappa_{1,t}$.

A.1.4 Solution of the numerical example

For the numerical example, we use the exact filtering equation, rather than the infinite recursion in theorem 1. To do so, we constrain the state x_t to lie on a discrete grid and time to increment in discrete steps. It then has a fixed set of transition probabilities. The component coming from the exponential jump we treat as geometrically distributed, while the component from the diffusion is single step up or down in the grid.

Figure A.1: Monthly estimates of the S&P 500 leverage effect



Note: The solid line is the time-series of coefficient estimates from regressions within each month of changes in volatility on returns on the SP 500. The shaded region is the 95-percent confidence interval each month.

x is therefore a Markov chain, and we treat the signal as $y_t \sim N\left(x_t, \sigma_y^2/\Delta t\right)$. Standard formulas then give the update for the posterior distribution over time. We calculate the VIX in the model as the instantaneous conditional volatility, using the formula from proposition 1.

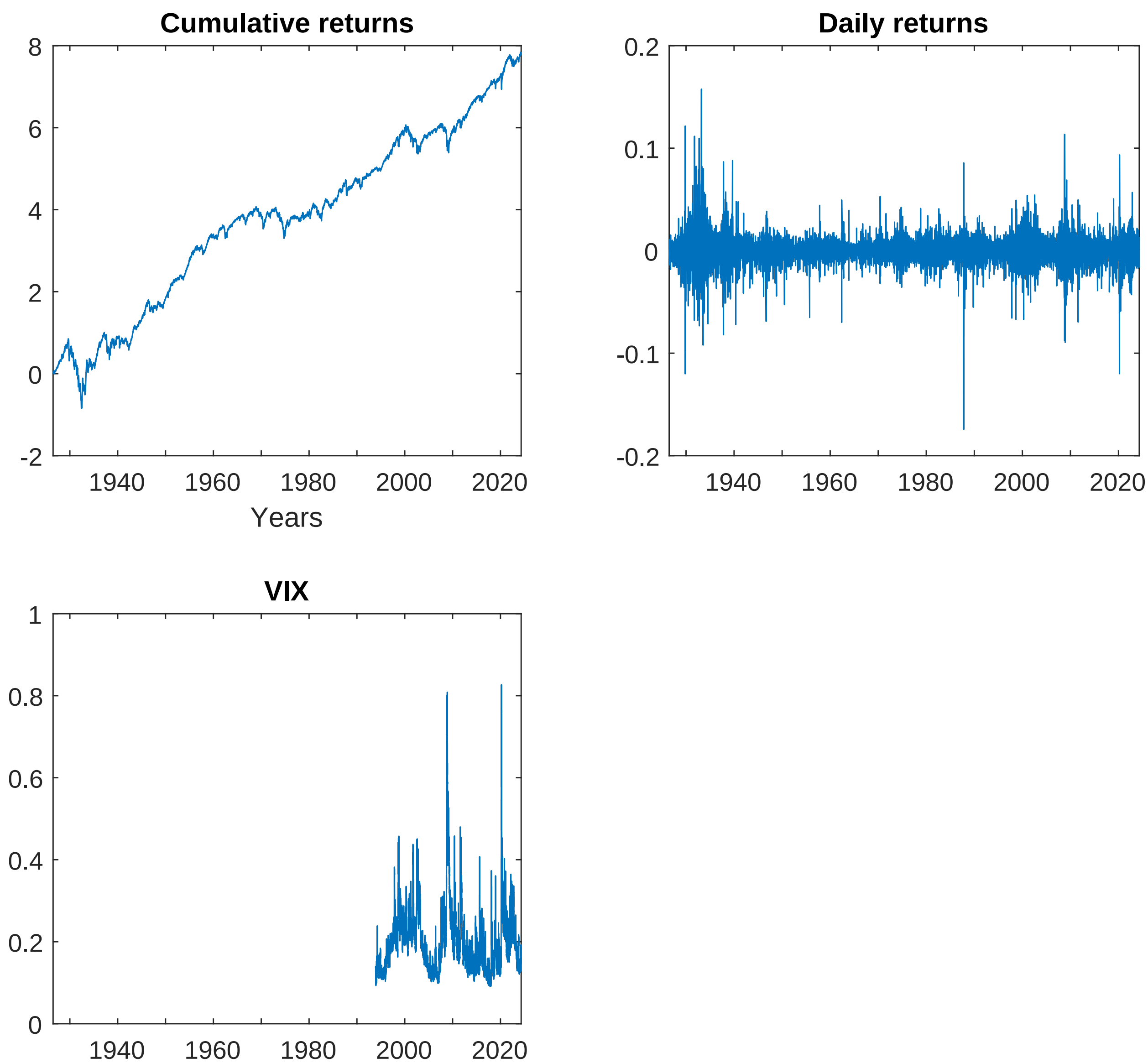
A.1.5 Moments for parameter selection

The first set of moments are unconditional moments of returns: the unconditional standard deviation and kurtosis and skewness at horizons of returns at one-, five-, 10-, and 20-day horizons.

The second is the same, but for returns scaled by lagged volatility, which we proxy for with the VIX. That is, we calculate the same unconditional moments for R_t/VIX_{t-1}.

The third set of moments is for daily changes in the VIX: their skewness, kurtosis, and correlation with market returns. Finally, the fourth set of moments is the 10-, 20-, and 60-day autocorrelations of the VIX.

Figure A.2: Empirical analogs to figure 5



Note: These plots report returns and the VIX in the data.