How beliefs respond to news: implications for asset prices

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

December 26, 2024

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 the expected discounted value of its future cash flows. In that sense, the stock market has over

*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, and Duke Fuqua.

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 dynamics of beliefs.[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 granularity when testing a model of beliefs.

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 NPV follows some arbitrary process, and agents continuously receive signals about it, which 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

[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).

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 *square*. That fact yields the sort of long memory that has been observed in stock market volatility. Rather than requiring a highly complicated model, or even any particular assumptions about fundamentals at all, long memory is an inevitable feature of information acquisition.[5]

Long memory also 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 should be related to the current conditional skewness of returns, and with a specific value for the coefficient, which we confirm in the data.

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

[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.

process is enough to generate 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 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 11 and 18 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 percent, which provides some independent support for our estimate (subject to the usual caveat that disagreement and uncertainty are theoretically distinct).

Stepping back

At a high level, the paper is about sufficient statistics. Without knowing the full model that agents believe drive fundamentals, we can still obtain strong implications for how beliefs are updated based on current beliefs. That basic fact has a surprising implication when it is reversed: by observing the local behavior of expectations – in our empirical setting, the local behavior of prices – it is possible to recover global features of beliefs, i.e. the conditional moments. Return volatility, even just within a single day, measures agents' uncertainty (and, again, we can put a number on that). The volatility of volatility measures the agents' conditional third moment.

On some level, that is inevitable – beliefs are the state variable, so they must determine the local behavior of prices. What is surprising, though, is that the relationships are universal, and not dependent on the underlying process, so it is possible to recover moments completely nonparametrically.

Past 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

[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.

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.

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.

Outline

The remainder of the paper is organized as follows. Section 2 describes the model structure and gives the main theoretical result. Section 3 then examines the theoretical predictions and section 4 studies some extensions and robustness to certain assumptions. Last, sections 5 and 6 take the model to the data, studying both a calibration and nonparametric tests of the theory, and section 7 concludes.

2 Model setup and solution

We motivate the analysis in terms of asset prices, and this section begins by describing the simple asset pricing framework we study. It is designed to lead to a standard filtering problem. The solution to that problem – in theorem 1 – is in fact a general result on Bayesian filtering, and not dependent on the asset pricing application.

2.1 Model setup

2.1.1 Dynamics of fundamentals

We assume 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 are (emphasizing their dependence on the information set)

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

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. 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, we represent the full set of information that an agent could possibly know by θ_t. The *fundamental value* of the asset is its price conditional on complete knowledge of θ_t,

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

While the class of admissible processes for D and M is very broad, it is not completely unconstrained. Instead of restricting them directly, though, we instead make assumptions about X itself, as it is the primary object of interest. A structural model will have implications for the dynamics of D and M, and one can then ask whether the induced X satisfies our assumptions. The necessary restrictions on X are reported as assumptions 1 and 2 in appendix A.1.1.1. The key requirement is that the pair $\{X, \theta\}$ is a Feller process (a slightly restricted version subset of the Markov processes). Examples of such processes are Brownian motions, jump diffusions, Levy-stable processes, and Itô semimartingales – in short, almost anything studied in economics. There is no requirement that X or θ is stationary at any order, and all of its moments need not exist (i.e. it may be heavy-tailed).

$\{X, \theta\}$ additionally can be the outcome of a learning model. For example, cash-flows might have a mean growth rate that is unknown, and then θ includes the entire history of cash-flows as part of the current state. Essentially what the Feller assumption rules out is just that the dynamics of $\{X, \theta\}$ are discontinuous in terms of the state. For example, the mean of X cannot shift discretely when the state θ crosses some boundary.[7]

2.1.2 Information flows

We assume that 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. $P_t(Y^t, \theta_t) = P_t(\theta_t)$), then we have, 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. We can use 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 in the case of inflation or interest rates), we could directly apply (3). To analyze stock prices, which are

[7] That might be a problem in certain models of bankruptcy, for example.
In principle, the results hold when θ_t is an element of a much more general class of topological spaces, and the Feller property is probably not strictly necessary. For example, under certain assumptions, the results here also hold when the state θ_t is itself a function (e.g. a cross-sectional distribution).

typically modeled as a geometric process, we restate the filtering problem in logs,

$$p_t \equiv E\left[x_t \mid Y^t\right] \tag{4}$$

$$\text{where } x_t = \log X_t \tag{5}$$

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 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.[8]

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.[9] 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 $\sigma_{Y,t} \to 0$.

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.[10] 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 4.2 discusses how the results apply when information may arrive discretely or when there are large 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 a pricing problem, but it is a much more general

[8]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.

[9]Formally, it is measurable with respect to the filtration induced by Y_t

[10]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.

setup. x is just some latent object of interest – it could be trend inflation, for example. Then $E\left[x_t \mid Y^t\right]$ would represent agents' expectations of trend inflation given their history of signals.

2.2 Solution to the filtering problem

We describe the dynamics of beliefs about fundamentals by describing the dynamics of the cumulants of agents' posteriors. Formally, the cumulants are the derivatives of the log characteristic function (properly factoring out powers of $\sqrt{-1}$),[11] but for the purposes of this paper we will only focus on the first three, which are equal (for all distributions for which they exist) to the first three central moments. In other words, almost nothing will be lost here if the word "moment" is substituted for "cumulant" in everything that follows.

Denote the n-th cumulant of the time-t conditional distribution of x_t by $\kappa_{n,t}$. Since the first cumulant is the expectation, $p_t = \kappa_{1,t}$.

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*[12]

$$d\kappa_{n,t} = \frac{\kappa_{n+1,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) - \frac{1}{\sigma_{Y,t}^2}\sum_{j=2}^{n}\alpha_j^{(n)}\kappa_{j,t}\kappa_{n-j+2,t}dt + E_t\left[d\left(x_t^k\right)\right] \tag{7}$$

where $E_t\left[\cdot\right] \equiv E\left[\cdot \mid Y^t\right]$ and the coefficients $\left\{\alpha_j^{(k)}\right\}$ are given in appendix A.1.1.

That result follows from a straightforward application of textbook results in Lipster and Shiryaev (2013) and Bain and Crisan (2009).[13] The most valuable feature of theorem 1 is that it shows that the sensitivity of each cumulant to signals is proportional to the current value of the next cumulant.

The first three cumulants, since they are equal to the first three moments, are worth writing out directly.

[11] The cumulants are sometimes defined as the derivatives of the cumulant generating function, $\log E\left[\exp\left(sx\right)\right]$, but that may not exist. The characteristic function always does, and when both functions exist, they have the same derivatives up to the $\sqrt{-1}$ scaling.

[12] 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.

[13] 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.

Corollary 2 *The dynamics of the first three moments/cumulants are*

$$dp_t = d\kappa_{1,t} = \frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[dx_t\right] \tag{8}$$

$$d\,\mathrm{var}_t\left[x_t\right] = d\kappa_{2,t} = \frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt \tag{9}$$

$$dE_t\left[\left(x_t - E_t\left[x_t\right]\right)^3\right] = d\kappa_{3,t} = \frac{\kappa_{4,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^3\right)\right] - \frac{3}{\sigma_{Y,t}^2}\kappa_{2,t}\kappa_{3,t}dt \tag{10}$$

The paper's predictions follow from these equations. Their key feature for our purposes is that the innovations in the cumulants are themselves multiplied by cumulants. The current cumulants are therefore sufficient statistics for their own dynamics, up to the $E_t\left[d\left(x_t^n\right)\right]$ terms, which depend on the dynamics of fundamentals.

The intuition for the result is simple: the filtering gain, $\kappa_{n+1,t}/\sigma_{Y,t}^2$, is a local regression coefficient. For the mean in equation (8), for example, the numerator $\kappa_{2,t}$ is equal to $\mathrm{cov}_t\left(x_t, dY_t\right)/dt$ and the denominator is equal to $\mathrm{var}_t\left(dY_t\right)/dt$, so their ratio is exactly the regression coefficient. Similarly, $\kappa_{3,t} = E\left[\left(x_t - E_t\left[x_t\right]\right)^3\right]$ 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 again a regression coefficient.

3 Predictions

We now examine the predictions of theorem 1 for the behavior of returns, beginning with volatility.

3.1 Volatility and the leverage effect

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 tiny compared to changes in prices.[14] We therefore treat return volatility as equal to price volatility. Formalizing the discussion above, we have

Corollary 3 *The instantaneous volatility of prices and hence returns is*

$$std\left(dp_t\right) = \frac{\kappa_{2,t}}{\sigma_{Y,t}}dt^{1/2} \tag{11}$$

[14] 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}.

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. If we can measure $\sigma_{Y,t}$ (which we show how to do below), then we can also measure investors' uncertainty about fundamentals, $\kappa_{2,t}$.

Proposition 1 *The instantaneous coefficient in a regression of changes in the conditional variance of returns on price changes is*

$$\frac{\operatorname{cov}\left(dp_t, d\left[std\left(dp_t\right)dt^{-1/2}\right]\right)}{\operatorname{var}\left(dp_t\right)} = \frac{\kappa_{3,t}}{\sigma_{Y,t}\kappa_{2,t}} \tag{12}$$

The presence of a **leverage effect** – a negative correlation between changes in volatility and prices – 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 (at least among those in our sample).

The intuition here 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. That intuition is generic – it is not dependent on the details of the specification. And, additionally, the value of theorem 1 is that it formalizes that intuition and shows that the third moment – as opposed to other concepts of asymmetry, like quantiles or higher-order moments – is in fact that correct measure to capture such an effect.

Figure A.1 plots monthly estimates of the leverage effect in the S&P 500 since 1990. The point estimate is negative in all but a single month in 1990 (and that value is not statistically significant), implying that investors have had negatively skewed conditional distributions for fundamentals consistently over the sample – in both good times and bad. 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 6.

3.2 Slow decay in volatility

The second term in (9) shows how volatility decays. When $\kappa_{2,t}$ (and hence also price volatility) is high, $\kappa_{2,t}^2\sigma_{Y,t}^{-2}dt$ 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.[15] When jumps up in $\kappa_{2,t}$ are large relative to its steady-state value, its decay is approximately of the form $1/(a+\Delta t)$ for a coefficient a (that depends on the other parameters of the model), and where Δt is the time since the shock. That is exactly the polynomial decay studied in the literature on long memory in volatility.

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 (9) shows it is actually a general feature of filtering problems.

The end result is that when investors are learning about fundamentals dynamically, long memory should be expected almost generically.[16] We examine the model's ability to fit detailed data on volatility dynamics in more detail in section 6.

3.3 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. Informally, to first order in $\sigma_{Y,t}(\Delta t)^{1/2}$,

$$skew_{t\to t+\Delta t}(dp_t) \approx 3\frac{\kappa_{3,t}}{\kappa_{2,t}}\sigma_{Y,t}(\Delta t)^{1/2} \tag{13}$$

$$= 3skew_t(x_t)\,\kappa_{2,t}^{1/2}\sigma_{Y,t} \tag{14}$$

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. But, locally, it scales with $skew_t(x_t)\,\kappa_{2,t}^{1/2}\sigma_{Y,t}$. That fact provides a link between indexes of the conditional skewness of returns ($skew_{t\to t+\Delta t}(dp_t)$), such as the CBOE's option-implied skewness, and

[15]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

[16]The "almost" here is because the effect can be quantitatively small. For small fluctuations, quadratic and linear mean reversion are indistinguishable.

the conditional skewness of fundamentals, $skew_t(x_t)$, which determines the leverage effect.

3.4 Skewness in volatility

In the data, the VIX is itself skewed. Table 1 in the quantitative analysis below shows that is true of both its level and monthly changes. The source of that effect is visible if we combine equations (9) and (13) (with $\Delta t \to dt$) to obtain

$$std\left(d\kappa_{2,t}\right) \approx \frac{\left|skew_t\left(dp_t\right)\right|}{3}\frac{\kappa_{2,t}}{\sigma^2_{Y,t}} \tag{15}$$

All else equal, the volatility of innovations to $\kappa_{2,t}$ scales with $\kappa_{2,t}$ itself. When $\kappa_{2,t}$ falls towards zero, the volatility of its innovations quickly becomes much smaller, while they grow when $\kappa_{2,t}$ rises. That effect creates a long right tail in the level of $\kappa_{2,t}$ itself, and any skewness in $\kappa_{2,t}$ itself is also inherited by $d\kappa_{2,t}$. 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).

3.5 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 $\kappa_{3,t} \neq 0$, long memory in volatility, skewness in returns for, skewness in both levels and changes in volatility, and time-varying volatility of volatility. None of this requires any mechanism more complicated than Bayesian updating in the presence of nonzero higher moments.

3.6 Examples

This section briefly considers two simple examples. Section 5 studies in much more depth a quantitatively realistic example.

3.6.1 Linear Gaussian process

If x is a linear Gaussian process (with a Gaussian prior at $t = 0$), then the model's solution is the Kalman filter. p 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 moments are always equal to zero. There is then no leverage effect, volatility of volatility, or skewness in expectations or volatility, as is well known.

As is well known, the Kalman filter, and hence the class of linear Gaussian processes, can accommodate parameter learning (Hamilton (1994)). Parameter learning can thus be accommodated within the dynamics for x, both in the Gaussian and non-Gaussian case.

3.6.2 Markov switching process

Veronesi (1999) studies a two-state switching model in which the latent state x_t 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 (see also Hamilton (1989)).

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:

$$\kappa_{2,t} = \pi_t (1 - \pi_t) \tag{16}$$

$$\kappa_{3,t} = (1 - 2\pi_t) \times \kappa_{2,t} \tag{17}$$

The variance here then is a bell-shaped function of π_t, peaking at 1/4 for $\pi_t = 1/2$, 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 near zero, the leverage effect *reverses*: agents no longer worry about the economy getting worse, so there is only 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.

4 Extensions and robustness

4.1 Inferring global properties of beliefs from local information

Theorem 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. This section reverses the analysis: how can we use local information about prices to learn about global properties of beliefs?

Under theorem 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. 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 theorem 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 $\kappa_{3,t}/\sigma_{Y,t}^2$.

4.2 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 2 *[Dytso, Poor, and Shamai (2022), equation (52)] For a random variable x_t and a signal $y_t \sim N\left(x_t, \sigma^2\right)$,*

$$\frac{d}{dy}\kappa_j\left(x_t \mid y_t = a\right) = \kappa_{j+1}\left(x_t \mid y_t = a\right) \tag{18}$$

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

Proposition 2 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 2 still requires normality.[17] That said, proposition 2 also shows why continuous time is useful here: the *prior* cumulants determine

[17]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 special cases, e.g. when the likelihood is exponential or Poisson.

sensitivities, rather than the *posterior* cumulants that appear in (18). In continuous time the cumulants follow continuous processes, so the prior and posterior values are essentially identical.

4.3 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 class of distributions 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 value 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. Those facts are also ways of seeing why there are no simple closed-form solutions to special cases of filtering problems outside the normal distribution.

5 Illustrative calibration

This section presents a simple quantitative example. We first use it to illustrate the model's core mechanisms by looking at impulse response functions, and then examine the extent to which the qualitative predictions above map into quantitatively reasonable behavior. The example demonstrates the extent to which 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 no 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.

5.1 Model setup and solution

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

$$dx_t = (\phi\lambda + g)\, dt - J_t dN_t \tag{19}$$

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.[18] 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 J_t 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 model – 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 a day.

5.2 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 $\{\phi, \lambda, \sigma_Y\} = \{0.00037, 0.43, 2.89\}$, where the time unit is taken to be a day. The value of ϕ implies that disasters occur on average once every 10.7 years, and λ implies that the average decline in fundamentals is 43 log points. The noise in signals, σ_y, is 289 log points per day. To get a sense of scale, if, hypothetically, agents had a prior variance for fundamentals of ∞ and fundamentals were constant, after one year of observing such signals their posterior standard deviation would be 0.11 in logs. We set $g = 0.07$ to match the historical equity premium.

5.3 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

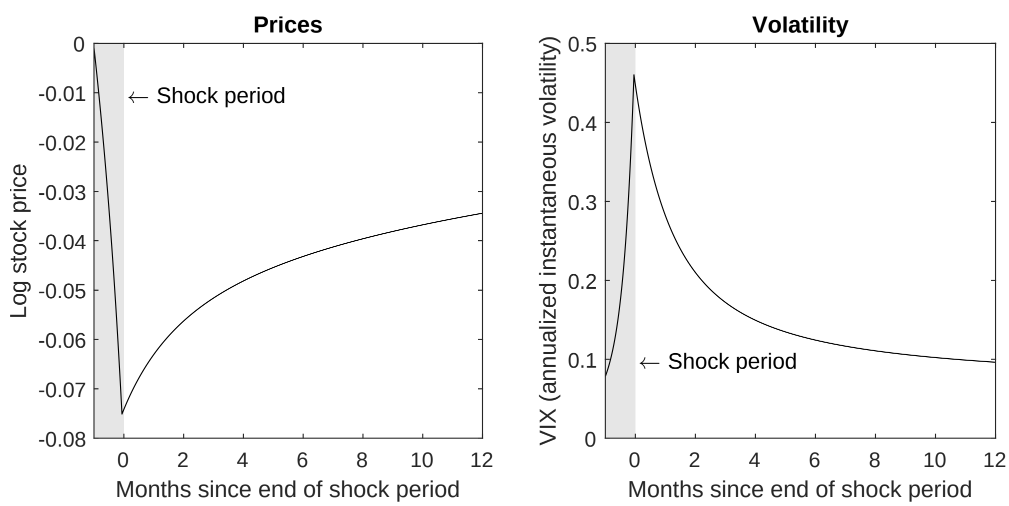
[18]The reduced-form process for x can easily be generated by assuming that cash-flows follow the same jump process. Positive risk premia can be generated by assuming the SDF is also driven by the same jumps (but with the opposite exposure).

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 2.4-standard deviation negative surprise (i.e. it is a 1-in-10 year event). 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 1 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.

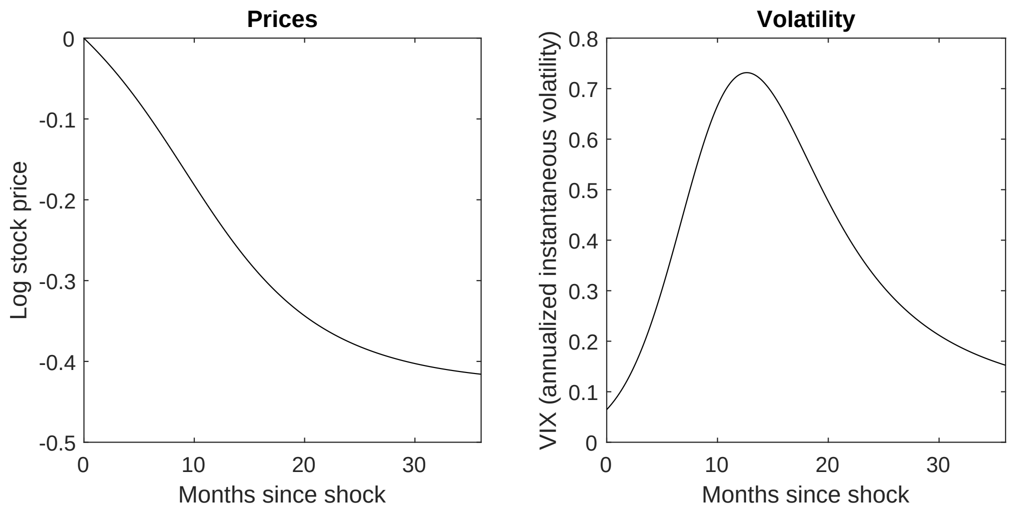
Figure 1: Response to negative error in the signal



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 $\sigma_Y dW$.

The second IRF is more standard – just a one-time unit-standard deviation (43 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 no noise, but, again, they still update beliefs as though there is noise).

Figure 2: 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.

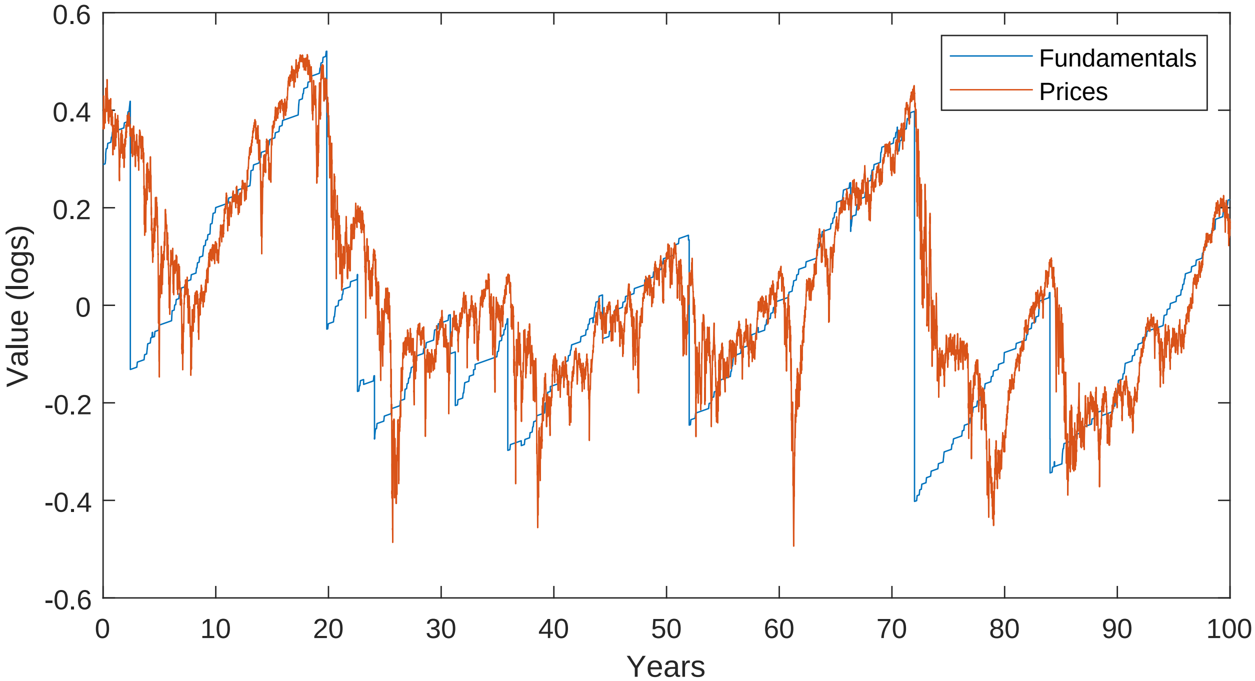
The left-hand plot in figure 2 shows that the decline in prices is again nonlinear – it accelerates before slowing, with the initial declines equal to 0.063 percent per day, accelerating to 0.1 percent per day at their peak. It takes about two years for prices to fully incorporate the drop in fundamentals. 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, which price declines accelerating and volatility accumulating over time.

5.4 Simulation results

The IRFs are in the end very stylized – they show responses to specific shocks and assuming a very particular path for the signals (and any single path is inevitably unlikely). Simulations are a better way of seeing what the full behavior of the model really is and how well it compares to the data.

Figure 3 plots the simulated time-series of fundamentals, x, and prices, p, 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.

Figure 3: 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.

We begin by examining unconditional moments for returns and the VIX in the model and the data in tables 1 and 2. Across the 13 moments, the model broadly matches the data, missing on only the most extreme statistics. Looking at returns, it has similar volatility and skewness, 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 returns scaled by lagged volatility, the model generates somewhat less kurtosis and skewness than is observed in the data, possibly indicating that it is failing to capture some intraday dynamics.

Table 2 shows that the model matches the data in terms of the unconditional standard deviation, skewness, and kurtosis for the VIX, both in levels and daily changes.

Table 1: Daily return moments

	R_t		R_t/VIX_{t-1}	
Moment	Data	Model	Data	Model
Std. dev.	1.49	1.21	1.00	1.00
Skewness	-0.16	-0.18	-0.58	-0.20
Kurtosis	19.4	10.4	5.5	3.1

Note: The table reports moments of the daily returns distribution, in the model and in the data.

Table 2: VIX moments

	Level		Daily change	
Moment	Data	Model	Data	Model
Std. dev.	7.9	7.4	1.6	1.1
Skewness	2.17	2.3	1.47	0.42
Kurtosis	11.41	10.4	29.9	21.2
Corr. w/ R_t	N/A	N/A	-0.70	-0.90

Note: The table reports moments of the VIX (level and daily changes), in the model and in the data.

Figure 4 plots the behavior of asset prices in a century of data simulated by the model. The shocks fed into the simulations are random, and not drawn to match the realized dynamics of stock returns over the past century. The simulations are just meant to evaluate whether the model's behavior *resembles* what is observed empirically.

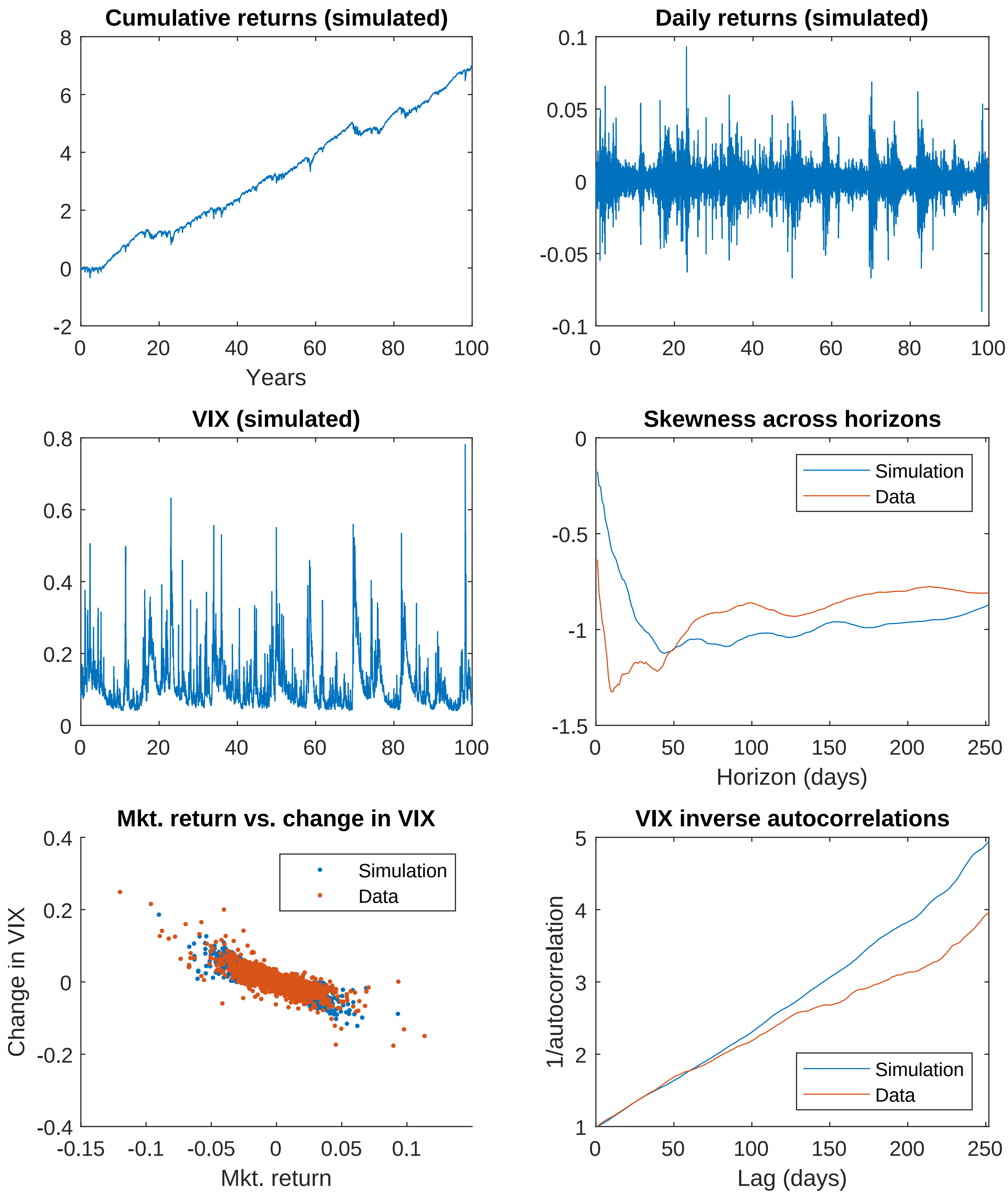
Panels (a), (b), and (c) show that the model generates stock returns and volatility that *appear*, both in levels and changes, highly similar to what is observed empirically. Mean stock returns here are set to zero, but any value can be achieved simply by modifying the dynamics of x – the model's testable predictions are about covariation among moments, rather than the mean of returns. The empirical analogs are reported in figure A.2 (though, again, what should match is the moments, not the specific realized paths of the variables).

Panel (d) shows that the behavior of skewness across horizons is also similar, though it accumulates somewhat more slowly in the model. Panels (e) and (f) report results for volatility dynamics. Panel (e) shows that the scatter plot of market returns against changes in the VIX is, in gross appearance, extremely similar in the model and data.

Panel (f) plots the multiplicative inverse of the autocorrelations of the VIX in the data and the model. Under the hypothesis that the autocorrelation in the VIX at lag j is approximately proportional to j^{-1}, the plot will be a straight line. On the other hand, if they decay geometrically, as they would (at least asymptotically) in an ARMA process (and at all lags in an AR(1), then the inverse autocorrelation curve will be convex. The values in the simulation and data are similar, and neither displays significant convexity, consistent with the presence of polynomial decay (and hence long memory).

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

Figure 4: Simulation results



Note: The plots are various outputs of a 100-year simulation of the model based on the parameters discussed in the text.

6 Estimated volatility dynamics and investor uncertainty

The analytic results in section 3 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.

6.1 Regression setup

Combining equations (11), (9), and (13), we have

$$d\left[std\left(dp_t\right)\right]=\frac{1}{\sigma_{Y,t}}E_t\left[d\left(x_t^2\right)\right]+\frac{1}{\sigma_{Y,t}^2}\left(\Delta t^{-1/2}skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right)dp_t-\frac{1}{\sigma_{Y,t}}\left[std\left(dp_t\right)\right]^2dt \tag{20}$$

If x has independent increments and σ_Y is constant – as in the quantitative model – then the first term is a constant.

What is particularly interesting about this regression is that it gives two separate estimates of the parameter σ_Y (when it is constant). The coefficient on $\left[std\left(dp_t\right)\right]^2$ depends on σ_Y since the average decline in uncertainty depends on the rate of information flow, and the coefficient on $skew \times dp$ depends on σ_Y^2 for the same reason. In addition to giving two estimates of a structural parameter, the regression therefore also has a testable implication, that the coefficient on $\left(\Delta t^{-1/2}skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right)dp_t$ should be the square of that on $\left[std\left(dp_t\right)\right]^2dt$.

For now we treat σ_Y as constant – the estimates can be thought of informally as measuring an average value of $\sigma_{Y,t}$. We relax that assumption below.

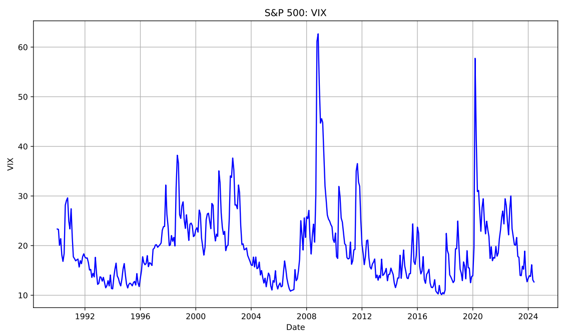
6.2 Data

We estimate the regression (20) 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 Cboe VIX index (rescaled consistent with the time units). For the return dp_t we use the log return on the CRSP total market index. Last, for $skew_t$, we use the Cboe SKEW index (properly transformed to correspond to a conditional skewness coefficient).[19] The $\Delta t^{-1/2}$ term in

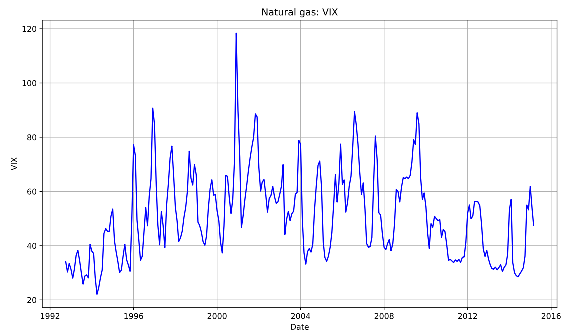
[19] That is, denoting the index value by $SKEW_t^{Cboe}$, we have that the skewness coefficient for returns is $skew_t\left(r\right)=\left(100-SKEW_t^{Cboe}\right)/10$.

parentheses in the middle term corresponds to the horizon at which skewness is calculated. We treat a unit of time as a day, so $\Delta t = 21$, representing 21 trading days, which is the approximate horizon of the Cboe SKEW index (one calendar month). We include the factor 3, so that the regressor is $\left(3 \times skew_t^{Cboe} \times 21\right)$, where $skew_t^{Cboe}$ is the transformed Cboe SKEW index. The top-left panel of figure 5 plots the Cboe VIX and the bottom-left implied skewness over our sample period. [20]

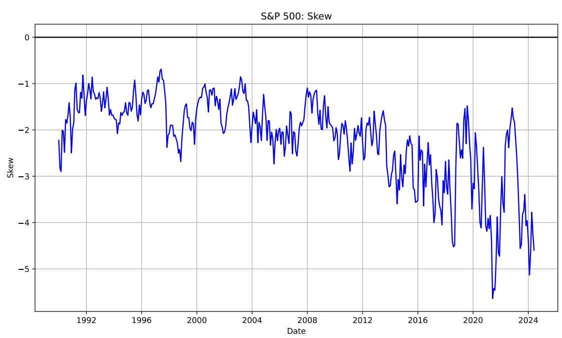
Figure 5: Time series of VIX and Skew for S&P 500 and natural gas



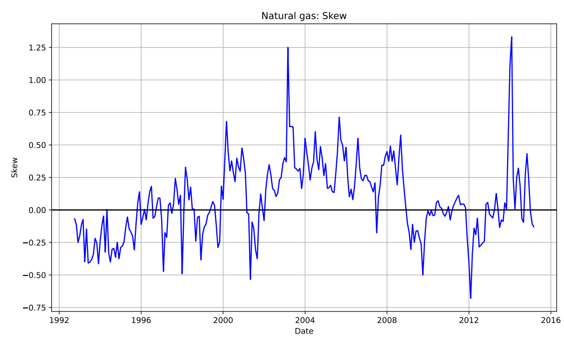
(a) VIX (S&P 500)



(b) VIX (natural gas)



(c) Skew (S&P 500)



(d) Skew (natural gas)

Note: Time series plots for the VIX and implied skewness for the S&P 500 and for natural gas. All series are monthly average of the corresponding daily series.

The S&P 500 implied skewness is almost exclusively negative, and returns are strongly negatively skewed. For the second market to use for estimation, we therefore looked for a large and economically significant market that displays positive skewness and settled on natural gas. For the underlying – i.e. to calculate dp_t – we use natural gas futures returns from the CME. We then use options on futures to calculate implied volatility and skewness using methods analogous to those used by the Cboe for the S&P 500 VIX and SKEW

[20] The Cboe skew index can be converted to an implied skew by subtracting 100 and multiplying by -0.1.

indexes.[21] The time series of the implied volatility and skewness for natural gas are plotted in the right-hand panels of figure 5. 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 should be much 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.

6.3 Results

Table 3 reports results of the regression implied by (20) 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 $\left(\Delta t^{-1/2} skew_{t \to t+\Delta t}(dp_t)\frac{1}{3}\right)$ should be the square of the coefficient on $std_{t-1}(dp_t)^2$. The bottom row of the table tests that hypothesis. For the S&P 500 the t-statistic is -1.05, while for natural gas it is -0.66, so the relative values of the coefficients are well within the range expected given statistical uncertainty.

In the data, $[std(dp_t)]^2$ is naturally fairly strongly correlated with $std(dp_t)$ itself, so one question is which dominates – the model says it should be $[std(dp_t)]^2$. The second and fourth columns of table 3 test that proposition by including both in the regression. $[std(dp_t)]^2$ does in fact appear to dominate. For the S&P 500, its t-statistic is larger by a factor of 4 than that for $std(dp_t)$, and it remains marginally statistically significant (at the 10-percent level). For natural gas, the coefficient on $std(dp_t)$ is statistically insignificant and economically indistinguishable from zero.

Similarly, we can ask whether it is $\left(\Delta t^{-1/2} skew_{t \to t+\Delta t}(dp_t)\frac{1}{3}\right) dp_t$ that dominates or simply the return, dp_t. The third and sixth columns of table 3 test that by including both variables. In this case, dp_t turns out to have a very slightly higher t-statistic for the S&P 500, but both variables remain individually highly significant. For natural gas,

We thus have a less strong confirmation of the model's prediction here – dp_t should have been driven out, but it is not. On the other hand, it still does not drive out the correct

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

Table 3: Volatility regressions

	S&P 500			Natural Gas		
	(1)	(2)	(3)	(4)	(5)	(6)
VIX^2_{t-1}	-0.56***	-0.68*	-0.49***	-0.87***	-1.01**	-0.81***
	[0.12]	[0.38]	[0.11]	[0.12]	[0.43]	[0.12]
$\frac{1}{3\sqrt{21}} skew_{t-1} dp_t$	0.46***	0.46***	0.22***	0.62***	0.62***	0.48***
	[0.01]	[0.01]	[0.02]	[0.05]	[0.05]	[0.05]
VIX_{t-1}		0.005			0.014	
		[0.010]			[0.032]	
dp_t			-0.039***			0.013***
			[0.004]			[0.001]
R^2	0.62	0.62	0.65	0.15	0.15	0.20
Test p-value	0.14			0.55		

Note: Daily regressions of first differences in the VIX for S&P 500 and daily natural gas volatility regressions. For the S&P, VIX is the CBOE 30-day VIX and Skew is the CBOE 30-day SKEW index (rescaled to correspond to 30-day implied skewness). For natural gas, VIX and Skew are option-implied quantities built from options data using methodologies paralleling the CBOE ones. Given that the natural gas contracts have varying maturities, we perform a panel regression controlling for contract fixed effects. Heteroskedasticity-robust standard errors are reported in brackets. Stars indicate statistical significance: * p¡.1, ** p¡.05, ***p¡.01.

variable, $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right) dp_t$. One possible explanation for this fact is that the Cboe SKEW index may be relatively noisy, since it is a higher-order moment, which would reduce the explanatory power of $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right) dp_t$ (and also bias its coefficient toward zero).

Overall the results show that the model's basic predictions seem to fit reasonably well, both for the S&P 500 and natural gas: we have approximately the right relationship between the coefficients on lagged implied variance and skewness$\times dp$, implied variance drives implied volatility out of the regression, and skewness$\times dp$ survives the inclusion of the two terms separately.

6.4 Evidence from volatility of volatility

There is a similar regression for the volatility of volatility. Again using (9) and (13) and combining now with (15), we have

$$std\left(d\left[std\left(dp_t\right)\right]\right) = \frac{1}{\sigma^2_{Y,t}} \times \frac{\left|skew_{t\to t+\Delta t}\left(dp_t\right)\right|}{3} \times std\left(dp_t\right) \tag{21}$$

This says that the volatility of volatility is proportional to volatility itself times conditional skewness. (21) now gives us a *third* estimate of σ_Y, and thus an additional testable prediction of the model.

For the S&P 500, volatility of volatility is available from the VVIX index, which is an implied volatility for the VIX itself based on options on VIX futures.[22] We again proxy for volatility and skewness using the CBOE indexes. There is no analog to the VVIX for natural gas (since there are not options on implied volatility for natural gas), so we only estimate this regression for the S&P 500.

Table 4 report results from the regression (21). The first column includes a constant – which is not estimated to be statistically significantly different from zero, consistent with the model – and the second column eliminates the constant. The coefficient estimate of about 0.30 is, again, an estimate of $1/\sigma_Y^2$, and consistent with the estimates from table 3. The R^2 in the regression ideally should be 1, and it is certainly not, but at the same time it is economically large. In addition to simple misspecification, deviations could come from measurement error in the asset prices or time variation in risk premia.

Table 4: Volatility-of-volatility regressions

	(1)	(2)
$\frac{1}{2} skew_t VIX_t$	0.30	0.32
	[0.05]	[0.02]
Constant	5.9×10^{-5}	N/A
	$[8.7\times10^{-5}]$	
R^2	0.47	0.85

Note: Daily regressions of the VVIX. Newey–West t-statistics are reported in brackets. Note that in the second column the R-squared is calculated based on the total sum of squares without demeaning, which is why it is much larger.

6.5 Comparing estimates of σ_Y for the S&P 500

So far we have three estimates of σ_Y – two from (20) and one from (21). There is also a fourth regression, related fairly closely to (20). In each month, we can estimate the leverage effect from a regression of changes in the VIX on market returns (see figure A.1). According to equation (12), if we then regress the estimated leverage effect in month t on implied skewness in month t (divided by 3), the coefficient is again an estimate of $1/\sigma_Y^2$.

Figure 6.5 compares the four different estimates of σ_Y for the S&P 500.

[22] Really, the VVIX measures the volatility of the log of the VIX. We therefore multiply the VVIX by the

Figure 6: Comparison of estimates of σ_Y

Note: Each dot is an estimate of σ_Y, and the whiskers represent 95

Estimate (1) is from the coefficient on VIX_{t-1} in equation (20); (2) is from the coefficient on lagged implied skewness times market returns in the same regression; (3) is from the volatility-of-volatility regression (21); and (4) is from the regression of the monthly leverage effect estimate on option-impleid skewness. The four estimates are all surprisingly consistent, ranging between about 1.5 and 1.8. Estimates (2)-(4) are all forms of the relationship between skewness, returns, and volatility. Estimate (1) is somewhat more independent, being based on the rate of mean reversion in volatility.

Part of what makes these estimates notable is that they are in a sense nonparametric: they do not require knowledge of the true dynamics of fundamentals. Additionally, the estimates are not completely inconsistent with the value of 2.89 used in the calibration (which was obtained from a moment-matching exercise). While the value is higher than those implied

VIX to get the volatility of the VIX in its own units.

from skewness and the leverage effect (estimates (2)-(4)), it is within the confidence band for the estimate based on mean reversion in volatility.

6.6 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

$$std\,(dp_t) = \frac{\kappa_{2,t}}{\sigma_{Y,t}} dt^{1/2} \tag{22}$$

$$\Rightarrow \kappa_{2,t}^{1/2} = \left(std\,(dp_t)\,\sigma_{Y,t} dt^{-1/2}\right)^{1/2} \tag{23}$$

Recall that the scaling of the estimates is for a unit time interval being equal to a day. The observed historical daily standard deviation of stock returns is about 1 percent. If $\sigma_{Y,t}$ is between 1.26 and 3.10 (based on the coefficient on VIX_{t-1} in equation (20), which is the most conservative of the confidence intervals), that implies that agents' posterior standard deviation is between 11.2 and 17.6 percent. The ± 2 standard deviation range for fundamentals around the current price for the aggregate stock market is then between ± 22.4 and ± 35.2 percent.

Similarly, we can get an estimate of average skewness in beliefs. One-month return skewness is historically approximately -2.2 (based on the average of the SKEW index). Plugging that into (13) 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 yet 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, so a survey giving a cross-section of estimates of fundamental value would be one way to validate our 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 fits inside the confidence band for uncertainty from equation (23) of [11.2,17.6]. That said, if we used the narrower confidence bands from the other estimates of σ_Y, the implied uncertainty would be somewhat lower.

6.7 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.

7 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

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

Bollerslev, Tim, 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.

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.

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.

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.

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

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.

Hamilton, James D, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica: Journal of the econometric society*, 1989, pp. 357–384.

Hamilton, James Douglas, *Time series analysis*, Princeton university press Princeton, 1994.

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

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.

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.

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

Revuz, Daniel and Marc Yor, *Continuous Martingales and Brownian Motion*, third ed., Vol. 293, Springer Science & Business Media, 1999.

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 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 $\{X, \theta\}$ *follows a Feller process with bounded and smooth functions in the domain of its extended infinitesimal generator.*[1]

[1] Feller processes are the subset of Markov processes for which the transition kernel varies continuously with the state. Throughout, we work with the càdlàg modification of x_t. For the definition of Feller processes and a proof of the existence of their càdlàg modification see Chapter 3, Section 2 of (Revuz and Yor 1999). If Z is a Feller process, a measurable function f is said to belong to the domain of the extended infinitesimal generator of Z if there exists a measurable function $\mathcal{G}_f$ such that, a.s., $\int_0^t |\mathcal{G}_f(Z_s)|ds < \infty$ for every t, and $f(Z_t) - f(Z_0) - \int_0^t \mathcal{G}_f(Z_s)ds$ is a right-continuous martingale for every initial state z ((Revuz and Yor 1999), pg. 285).

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

Assumption 3 *For all* t*, the noise volatility satisfies*

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

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

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

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

where Y *and* $\tilde{Y}$ *are two different realizations of the signal process,* L_1 *and* L_2 *are non-negative constants, and* $K(t)$ *is a non-decreasing right-continuous function satisfying* $0 \le K(t) \le 1$ *for all* t*.*

A.1.1.2 Proof

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

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

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

Proof. The lemma obtains as an application of Theorem 8.1 of (Liptser and Shiryaev 2013) by setting $h_t \to \exp(i\omega x_t)$, $\xi_t \to Y_t$, $A_t \to x_t$, and $B_t(\xi) \to \sigma_{Y,t}$.[2] We proceed by verifying that conditions (8.1)–(8.9) of (Liptser and Shiryaev 2013) are satisfied.

Equation (8.2) is simply equation (2) of the paper in integral form. Since $f(x) \equiv \exp(i\omega x)$ is a smooth and bounded function, by Assumption 1, there exists a measurable function $\mathcal{G}_f$ such that, almost surely, $\int_0^t |\mathcal{G}_f(x_s)|ds < \infty$ for every t and $M_t \equiv f(x_t) - f(x_0) - \int_0^t \mathcal{G}_f(x_s)ds$ is a right-continuous martingale. Therefore, condition (8.1) is satisfied. Furthermore, since f is a bounded function, so is $\mathcal{G}_f$. Consequently, conditions (8.6) and (8.7) are also satisfied. Conditions (8.4), (8.5), (8.9), and the second part of condition (8.3) are satisfied

[2]The results cited here are stated for real-valued functions. However, they 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.

by Assumption 3. Finally, condition (8.8) and the first part of condition (8.3) are satisfied by Assumption 2. Applying Theorem 8.1 and noting that the Brownian motion W_t is independent of x_t, we get

$$\mathbb{E}_t[\exp(i\omega x_t)] = \mathbb{E}_0[\exp(i\omega x_0)] + \int_0^t \mathbb{E}_s[\mathcal{G}_f(x_s)]ds + \int_0^t \frac{\operatorname{cov}_s(x_s, \exp(i\omega x_s))}{\sigma_{Y,s}} d\overline{W}_s, \quad \text{(A.5)}$$

where

$$\overline{W}_t = \int_0^t \frac{dY_s - \mathbb{E}_s[x_s]ds}{\sigma_{Y,s}}. \quad \text{(A.6)}$$

Or equivalently,

$$d\mathbb{E}_t[\exp(i\omega x_t)] = \mathbb{E}_t[\mathcal{G}_f(x_t)]dt + \operatorname{cov}_t(x_t, \exp(i\omega x_t)) \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_{Y,t}^2}. \quad \text{(A.7)}$$

On the other hand, by the definition of $\mathcal{G}_f$,

$$d\exp(i\omega x_t) - \mathcal{G}_f(x_t)dt = dM_t, \quad \text{(A.8)}$$

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

Theorem 1 *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 the nth moment of x_t exist, and Assumptions 1–3 are satisfied. Then for every $k \leq n$,*

$$d\kappa_{k,t} = \mathbb{E}_t[d(x_t^k)] + \frac{\kappa_{k+1,t}}{\sigma_{Y,t}^2}(dY_t - \mathbb{E}_t[x_t]dt) - \frac{1}{\sigma_{Y,t}^2}\sum_{j=2}^{k} \alpha_j^{(k)} \kappa_{j,t}\kappa_{k-j+2,t}dt, \quad \text{(A.9)}$$

where $\alpha_j^{(k)}$ are constants, defined recursively as follows: $\alpha_2^{(2)} = 1$, and

$$\alpha_j^{(k+1)} = \begin{cases} 1 + \alpha_j^{(k)} \text{ if } j = 2 \\ \alpha_{j-1}^{(k)} + \alpha_j^{(k)} \text{ if } 3 \leq j \leq k \\ \alpha_{j-1}^{(k)} \text{ if } j = k+1 \end{cases} \quad \text{(A.10)}$$

Proof. By Lemma 1 and Itô's lemma,

$$d\log\varphi_{x,t}(\omega) = \frac{\mathbb{E}_t[d\exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} + \frac{\operatorname{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_{Y,t}^2} - \frac{1}{2\sigma_{Y,t}^2}\left(\frac{\operatorname{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\right)^2 dt. \quad \text{(A.11)}$$

Since the posterior distribution of x_t has $n+1$ moments, it also has $n+1$ cumulants and the corresponding characteristic function has $n+1$ derivatives at $\omega = 0$, where the cumulants are related to the derivatives of the characteristic function via[3]

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

The key step is to differentiate equation (A.11). Differentiating the left-hand side with respect to ω and applying the dominated convergence theorem we get

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

For the first term on the right-hand side of (A.11), 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,

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

for any $k \leq n+1$. On the other hand,

$$\mathbb{E}_t[d \exp(i\omega x_t)] \big|_{\omega=0} = 0, \tag{A.16}$$

$$\mathbb{E}_t[\exp(i\omega x_t)] \big|_{\omega=0} = 1. \tag{A.17}$$

Consequently, for all $k \leq n+1$,

$$\frac{d^k}{d\omega^k} \frac{\mathbb{E}_t[d \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} \Bigg|_{\omega=0} \tag{A.18}$$

$$= \frac{1}{\mathbb{E}_t[\exp(i\omega x_t)]} \frac{d^k}{d\omega^k} \mathbb{E}_t[d \exp(i\omega x_t)] \Bigg|_{\omega=0} + \mathbb{E}_t[d \exp(i\omega x_t)] \frac{d^k}{d\omega^k} \left(\mathbb{E}_t[\exp(i\omega x_t)] \right)^{-1} \Bigg|_{\omega=0} \tag{A.19}$$

$$= i^k \mathbb{E}_t[d(x_t^k)]. \tag{A.20}$$

[3] 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 in (A.11), note that

$$\frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} = \frac{\mathbb{E}_t[x_t \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} - \mathbb{E}_t[x_t] = i^{-1}\frac{d}{d\omega}\log\varphi_{x,t}(\omega) - \mathbb{E}_t[x_t]. \tag{A.21}$$

Therefore,

$$\frac{d^k}{d\omega^k}\frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0}\frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma^2_{Y,t}} = i^{-1}\frac{d^{k+1}}{d\omega^{k+1}}\log\varphi_{x,t}(\omega)\bigg|_{\omega=0}\frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma^2_{Y,t}} \tag{A.22}$$

$$= i^k\frac{\kappa_{k+1,t}}{\sigma^2_{Y,t}}\left(dY_t - \mathbb{E}_t[x_t]dt\right). \tag{A.23}$$

Finally, we compute the derivative of the last term in (A.11) using induction. To simplify the notation, let $f'^{-1}\frac{d}{d\omega}\log\varphi_{x,t}(\omega)$ and $c \equiv \mathbb{E}_t[x_t]$. We are interested in the kth derivative of $\frac{1}{2}\left(f'(\omega) - c\right)^2$ evaluated at $\omega = 0$. In what follows, we first prove by induction that

$$\frac{d^k}{d\omega^k}\frac{\left(f'(\omega) - c\right)^2}{2} = \left(f'(\omega) - c\right)f'^{(k+1)}(\omega) + \sum_{j=2}^{k}\alpha_j^{(k)}f^{(j)}(\omega)f^{(k-j+2)}(\omega), \tag{A.24}$$

where constants $\alpha_j^{(k+1)}$ are as in the statement of the theorem. Note that

$$\frac{d}{d\omega}\frac{\left(f'(\omega) - c\right)^2}{2} = (f'(\omega) - c)f''(\omega), \tag{A.25}$$

$$\frac{d^2}{d\omega^2}\frac{\left(f'(\omega) - c\right)^2}{2} = (f'(\omega) - c)f'''(\omega) + (f''^2. \tag{A.26}$$

Therefore, the induction base holds with

$$\alpha_2^{(2)} = 1.$$

Now suppose the induction hypothesis holds for k. Then,

$$\frac{d^{k+1}}{d\omega^{k+1}}\frac{\left(f'\left(\omega\right)-c\right)^2}{2} = \left(f'\left(\omega\right)-c\right) f'^{(k+2)}(\omega) + f^{(2)}(\omega) f^{(k+1)}(\omega) \tag{A.27}$$

$$+ \sum_{j=2}^{k} \alpha_j^{(k)} f^{(j+1)}(\omega) f^{(k-j+2)}(\omega) + \sum_{j=2}^{k} \alpha_j^{(k)} f^{(j)}(\omega) f^{(k-j+3)}(\omega) \tag{A.28}$$

$$= \left(f'\left(\omega\right)-c\right) f'^{(k+2)}(\omega) + f^{(2)}(\omega) f^{(k+1)}(\omega) \tag{A.29}$$

$$+ \sum_{j=3}^{k+1} \alpha_{j-1}^{(k)} f^{(j)}(\omega) f^{(k-j+3)}(\omega) + \sum_{j=2}^{k} \alpha_j^{(k)} f^{(j)}(\omega) f^{(k-j+3)}(\omega) \tag{A.30}$$

$$= \left(f'\left(\omega\right)-c\right) f'^{(k+2)}(\omega) + \sum_{j=2}^{k} \alpha_j^{(k+1)} f^{(j)}(\omega) f^{(k-j+3)}(\omega), \tag{A.31}$$

where $\alpha_j^{(k+1)}$ is given by (A.10). Noting that $f'(0) = c$ and $i^{-k+1} f^{(k)}(0) = \kappa_{k,t}$, we get the following expression for the derivative of the last term in (A.11):

$$\frac{d^k}{d\omega^k}\frac{1}{2\sigma_{Y,t}^2}\left(\frac{\operatorname{cov}_t(\exp(i\omega x_t), x_t)}{\mathbb{E}_t[\exp(i\omega x_t)]}\right)^2 dt\bigg|_{\omega=0} = \frac{i^k}{\sigma_{Y,t}^2}\sum_{j=2}^{k}\alpha_j^{(k)}\kappa_{j,t}\kappa_{k-j+2,t}dt. \tag{A.32}$$

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

Corollary 1 *Suppose the Assumptions of Theorem 1 are satisfied, and additionally x_t is a martingale. Then,*

$$dp_t = \frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t[x_t]dt\right), \tag{A.33}$$

$$d\operatorname{var}_t[x_t] = \frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t[x_t]dt\right) - \frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt + \mathbb{E}_t[(dx_t)^2], \tag{A.34}$$

$$dE_t\left[\left(x_t - E_t\left[x_t\right]\right)^3\right] = \frac{\kappa_{4,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t[x_t]dt\right) - \frac{3\kappa_{2,t}\kappa_{3,t}}{\sigma_{Y,t}^2}dt + \mathbb{E}_t[(dx_t)^3] + 3\operatorname{cov}_t(x_t,(dx_t)^2). \tag{A.35}$$

Corollary 1 *Suppose the Assumptions of Theorem 1 are satisfied, and additionally x_t has* *independent increments. Then,*

$$dp_t = \frac{\kappa_{2,t}}{\sigma^2_{Y,t}} \left(dY_t - \mathbb{E}_t[x_t]dt\right) + \mathbb{E}_t[dx_t], \tag{A.36}$$

$$d\,\mathrm{var}_t[x_t] = \frac{\kappa_{3,t}}{\sigma^2_{Y,t}} \left(dY_t - \mathbb{E}_t[x_t]dt\right) - \frac{\kappa^2_{2,t}}{\sigma^2_{Y,t}}dt + \mathbb{E}_t[(dx_t)^2], \tag{A.37}$$

$$dE_t\left[(x_t - E_t\left[x_t\right])^3\right] = \frac{\kappa_{4,t}}{\sigma^2_{Y,t}} \left(dY_t - \mathbb{E}_t[x_t]dt\right) - \frac{3\kappa_{2,t}\kappa_{3,t}}{\sigma^2_{Y,t}}dt + \mathbb{E}_t[(dx_t)^3]. \tag{A.38}$$

A.1.2 Derivation of equation (13)

We calculate the local skewness over a horizon of length Δt by informally integrating the differential equation for the first cumulant. Specifically,

$$\kappa_{1,t+\Delta t} \approx \kappa_{1,t} + \frac{\kappa_{2,t}}{\sigma^2_t}\hat{Y}_{t+\Delta t} + \frac{1}{2}\frac{\kappa_{3,t}}{\sigma^2_t}\left(\hat{Y}^2_{t+\Delta t} - \sigma^2_t \Delta t\right) \tag{A.39}$$

$$\text{where } \hat{Y}_{t+\Delta t} \equiv Y_{t+\Delta t} - Y_t - E_t x_t \Delta t \tag{A.40}$$

We then have that

$$E_t\left[\kappa_{1,t+\Delta t}\right] = \kappa_{1,t} \tag{A.41}$$

$$E_t\left[\left(\kappa_{1,t+\Delta t} - E_t\kappa_{1,t+\Delta t}\right)^2\right] = \kappa^2_{2,t}\sigma_t^{-2}\Delta t + o\left(\Delta t\right) \tag{A.42}$$

$$E\left[\left(\kappa_{1,t+\Delta t} - E_t\kappa_{1,t+\Delta t}\right)^3\right] = 3\kappa^2_{2,t}\kappa_{3,t}\sigma_t^{-2}\Delta t^2 + o\left(\Delta t^2\right) \tag{A.43}$$

So the skewness coefficient is

$$\frac{E\left[\left(\kappa_{1,t+\Delta t} - E_t\kappa_{1,t+\Delta t}\right)^3\right]}{E_t\left[\left(\kappa_{1,t+\Delta t} - E_t\kappa_{1,t+\Delta t}\right)^2\right]^{3/2}} \approx \frac{3\kappa^2_{2,t}\kappa_{3,t}\sigma_t^{-2}\Delta t^2}{\left(\kappa^2_{2,t}\sigma_t^{-2}\Delta t\right)^{3/2}} \tag{A.44}$$

$$= 3\kappa^{-1}_{2.t}\kappa_{3,t}\sigma_t\left(\Delta t\right)^{1/2} \tag{A.45}$$

A.1.3 Derivation of equation (20)

Starting from equation (9),

$$
\begin{aligned}
d\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right) &= \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt && \text{(A.46)}\\
&= \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\kappa_{2,t}}\frac{\kappa_{2,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt && \text{(A.47)}\\
&= \frac{1}{\sigma_{Y,t}^{2}}\left(skew_{t\to t+\Delta t}\left(dp_t\right)\left(\Delta t\right)^{-1/2}\right)\frac{1}{3}\frac{\kappa_{2,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt && \text{(A.48)}\\
&= \frac{1}{\sigma_{Y,t}^{2}}\left(skew_{t\to t+\Delta t}\left(dp_t\right)\left(\Delta t\right)^{-1/2}\right)\frac{1}{3}\left[dp_t - E_t dx_t\right] + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt && \text{(A.49)}
\end{aligned}
$$

where the third line uses equation (13) and the fourth line inserts the formula for $dp_t = d\kappa_{1,t}$.

A.1.4 Exponential model solution

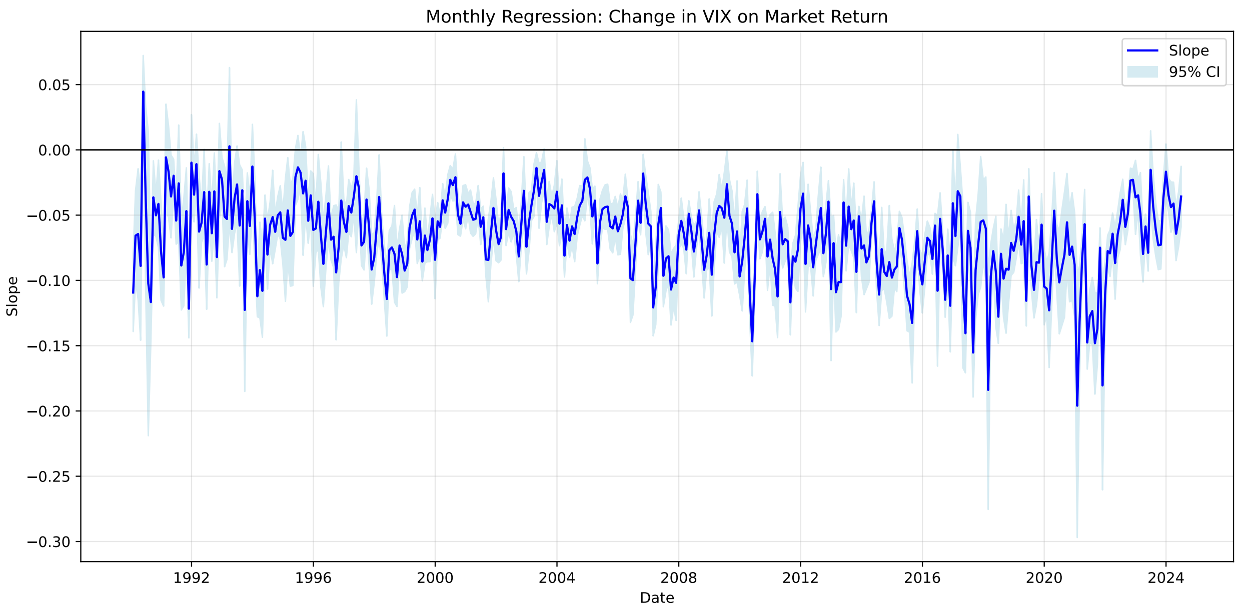
A.1.5 Moments for parameter selection

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}.

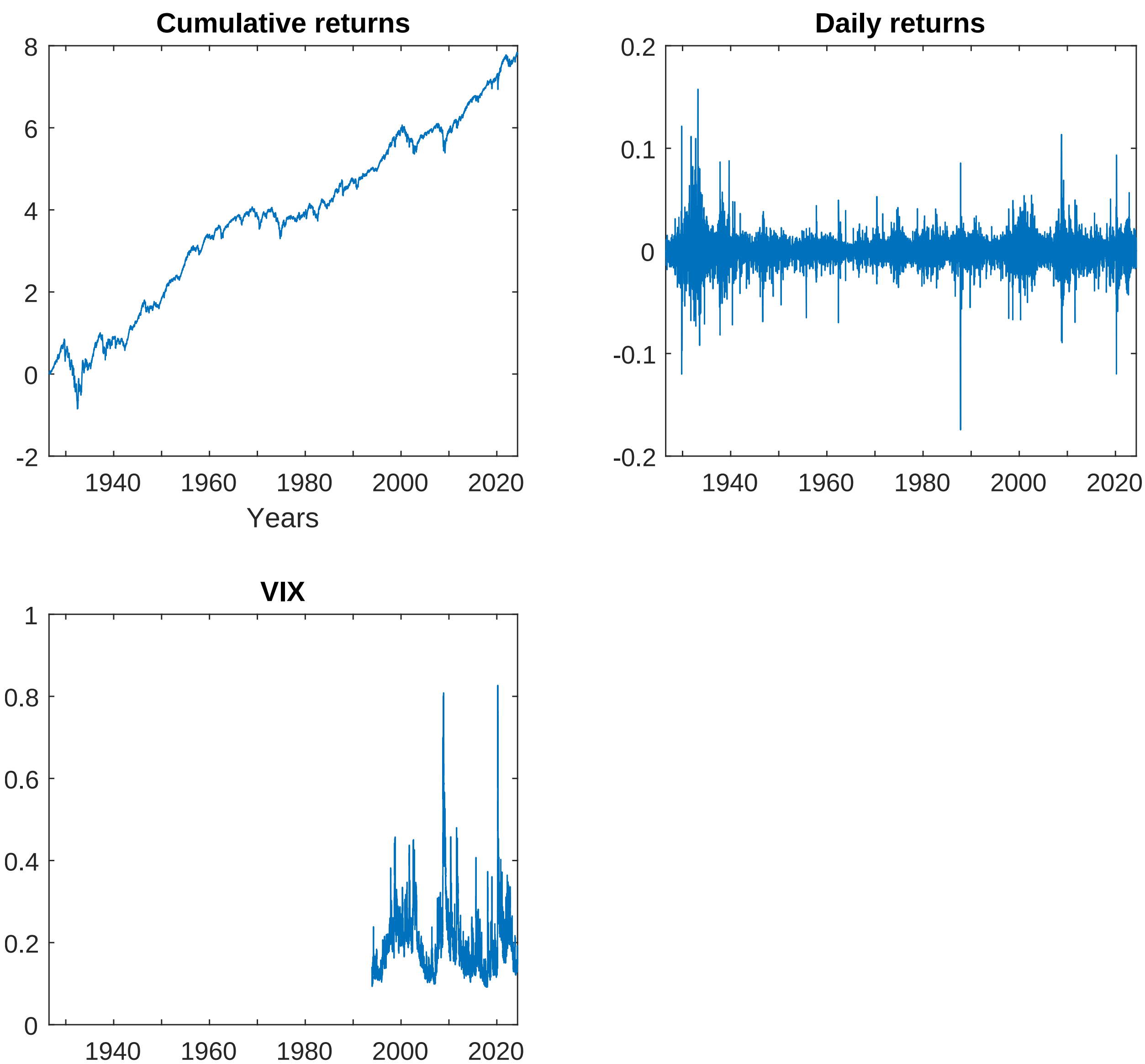
The third set of moments is for daily changes in the VIX: their skewness, kurtosis, and correlation with market returns. Finally, the fourth set of moments is the 10-, 20-, and 60-day autocorrelations of the VIX.

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



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

Figure A.2: Empirical analogs to figure 4



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