The Inherent Nonlinearity in Learning:

Implications for Understanding Stock Returns

Ian Dew-Becker, Stefano Giglio,
and Pooya Molavi

August 27, 2025

WP 2025-16

https://doi.org/10.21033/wp-2025-16

FEDERAL RESERVE BANK *of* CHICAGO

The inherent nonlinearity in learning: implications for understanding stock returns

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

August 27, 2025

Abstract

Financial markets (and more generally the real economy) display a wide range of important nonlinearities. This paper focuses on stock returns, which are skewed left – generating crashes – and have volatility that moves over time, is itself skewed, is strongly related to the level of prices, and displays long memory. This paper shows that such behavior is actually almost inevitable when prices are formed by investors acquiring information about the true, but latent, value of stocks. It studies a general model of filtering in which agents receive signals about the fundamental value of the stock market and dynamically update their beliefs (potentially with biases). When those beliefs are non-normal and investors believe crashes can happen, prices generically display the range of nonlinearities observed in the data. While the model does not explain where crashes come from, it shows that investors believing that prices can crash is sufficient to generate the rich higher-order dynamics observed empirically. In a simple calibration with iid shocks to fundamentals, the model fits well quantitatively, and regression-based tests support the model's mechanism.

1 Introduction

Background and contribution

*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 and discussions from Sergei Glebkin, Jiantao Huang, Paymon Khorrami, Jean-Paul Renne, Larry Schmidt, and seminar participants at the Bank of Canada, Duke, Stanford, Indiana, Yale, the San Francisco and Chicago Feds, and the Transatlantic Theory, CEPR Beliefs and the Macroeconomy, INSEAD, FIRS, HKU Macro, and Saieh Fellows conferences.

Stock market returns are far from normally distributed. The most salient deviation is that the market sometimes crashes, so that returns are skewed left. Crashes and negative skewness more generally have been the subject of huge amounts of research trying to understand both their causes – perhaps the bursting of bubbles, for example – and also their consequences. Do stock market crashes cause declines in GDP or does the causation run the opposite direction? Does crash risk explain the equity premium?[1]

But stock market returns are much more complicated than simply being independent draws from a negatively skewed distribution. Their volatility fluctuates over time, and those fluctuations have also been the subject of large literatures in both macroeconomics and finance.[2] Movements in volatility are themselves positively skewed and they have a strong negative correlation with market returns (about -80%) known as the **leverage effect**.[3] The strength of that relationship also changes over time and is related to the conditional skewness of market returns (Neuberger (2012)), which itself moves over time and is related to the state of the macroeconomy.[4] Finally, volatility has nonlinear dynamics: following large increases, instead of decaying geometrically, it tends to revert relatively quickly in the short-run and then more slowly.[5]

The basic contribution of this paper is to show that not only is it not surprising that the stock market behaves in those ways, but that such behavior is nearly inevitable when its level is set by investors who are continuously trying to learn about the true value of stocks.[6] While there is work that has studied different aspects of stock return nonlinearity individually, this paper is the first to propose a joint explanation. Its analysis cannot explain why there are crashes; instead what it shows is that in a world where crashes happen and investors are continually acquiring information, those crashes should happen in a consistent way: with

[1]On bubbles and crashes, among many others, see Abreu and Brunnermeier (2003) and Phillips, Shi and Yu (2015). For their relationship with GDP, see Reinhart and Rogoff (2009) and Sufi and Taylor (2022). On crash risk and the equity premium, see Rietz (1988) and Barro (2006).

[2]E.g. Bloom (2009) studying shocks to the VIX and the subsequent literature that follows. More recently, see Caldara, Fuentes-Albero, Gilchrist and Zakrajšek (2016) and Ludvigson, Ma and Ng (2021).

[3]See Merton (1980), French, Schwert and Stambaugh (1987), and Cont (2001). The term is generally viewed as a misnomer – while financial leverage can qualitatively generate countercyclical volatility, the effect would be smaller than what is observed empirically by an order of magnitude.

[4]For recent work, see Salgado, Guvenen and Bloom (2020), Gormsen and Jensen (2023), Iseringhausen, Petrella and Theodoridis (2023), Dew-Becker (2024), and Menkhoff (2025).

[5]E.g. Mandelbrot (1963), Granger (1980), and Mandelbrot, Fisher and Calvet (1997). Cont (2001) gives a thorough and still relevant review of the facts for volatility dynamics and other nonlinearities in financial markets.

[6]For past work on belief dynamics and stock returns, among many others, see David (1997), Veronesi (1999), Weitzman (2007), David and Veronesi (2013), Collin-Dufresne, Johannes and Lochstoer (2016), Gennaioli, Shleifer and Vishny (2015), Johannes, Lochstoer and Mou (2016), Kozlowski, Veldkamp and Venkateswaran (2018), Farmer, Nakamura and Steinsson (2024), Wachter and Zhu (2023), and Orlik and Veldkamp (2024).

volatility rising as prices fall and then nonlinearly returning to its mean, and the strength of that relationship being related to the conditional skewness of returns. Information processing naturally and almost inevitably creates the pattern of nonlinearities observed in asset prices.

Methods

The paper's theoretical structure is built around the premise that agents want to know the discounted value of a security's cash flows. It uses a very general setup: the net present value (NPV) follows some arbitrary process, and agents continuously receive signals about it. Since the NPV process is essentially unconstrained, the analysis nests a wide range of specifications that have been studied in the literature.

Asset prices in the model are the solution to a filtering problem: agents observe signals about the true value of stocks and prices depend on their posterior mean. At any given time, investors in fact have not just a mean but a full posterior distribution over the possible fundamental value of the stock market.

The paper's core theoretical tool is a novel result showing that belief dynamics have a simple recursive structure: the sensitivity of the posterior mean to signals is equal to the posterior variance multiplied by signal precision, and the sensitivity of the posterior variance is equal to the posterior third moment times signal precision. The result also yields an expression for the mean reversion in the posterior variance. In fact nothing about that is restricted to an asset pricing setting; it is a much more general statement about the dynamics of beliefs. The methods also do not require full rationality – agents' beliefs about crash risk, for example, could be misspecified.

Results

The first important feature of the theoretical results is that they immediately imply there is a tight relationship between investors' uncertainty about fundamentals and return volatility – high uncertainty in the model creates high return volatility, since uncertainty causes agents to respond strongly to the signals they observe. That then helps understand the leverage effect – the increase in volatility as prices fall – which is possibly the strongest and most consistent of the nonlinearities in the data. There is a necessary and sufficient condition for the leverage effect to appear in the model: agents must have negatively skewed beliefs about fundamentals. When agents' subjective distribution over the true value for fundamentals is negatively skewed, a negative signal – which drives their mean, and hence prices, down – also raises investor uncertainty, since negative skewness means that the left-hand side of their distribution is wider than the right. So when investors have negatively skewed beliefs, negative news both reduces prices and raises uncertainty and therefore price volatility.

It is again important to note that the paper has nothing to say about why investors have negatively skewed beliefs. Naturally that skewness may come from the facts that crashes do happen and that other features of the economy also display negative skewness.[7] But behavioral biases or a type of ambiguity aversion could also play a role.

Unsurprisingly, the strength of the leverage effect in the model is related to the magnitude of skewness in beliefs. In a simple empirical analysis looking at the US stock market and natural gas futures (with the latter selected for having strongly positive skewness in contrast to the stock market), the leverage effect coefficient lines up strikingly well with the model's prediction.

The paper next shows that nonlinear filtering also generically yields long memory in volatility, i.e. nonlinear decay following shocks. When uncertainty is high, agents put high weight on the signals they receive, causing them to learn and reduce uncertainty quickly. But as uncertainty falls, they also naturally give each additional signal less weight, causing their learning to slow. The result is that following upward jumps, the rate of mean reversion is high initially and then slows, consistent with the data.

After developing a few more theoretical results, the paper moves on to a quantitative analysis that examines the model's predictions from two perspectives. First, since the theoretical results are primarily qualitative and in many cases rely on certain limits (e.g. very small time periods), we examine a simple calibration to see whether the model's mechanisms are quantitatively realistic. The calibration is set up so that the latent value of stocks follows an i.i.d. disaster process with jump sizes distributed as in the estimates of Barro and Jin (2011) for global consumption disasters, while the signal precision is taken as a free parameter to match the data. Importantly, because fundamentals follow a random walk in the calibration, without the learning there would be no nonlinearities in returns other than skewness.

The calibration provides a good fit to 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: that the filtering mechanism is potentially quantitatively relevant and that the addition of an extremely simple learning process to a standard disaster setup (which by itself would have disasters but none of the other real-world nonlinearities) generates quantitatively significant nonlinearity – enough to match what is observed in stock return dynamics, even in samples in which there are no large disasters.[8]

[7] For empirical work, see Sichel (1993), McKay and Reis (2008), Morley and Piger (2012), Berger, Dew-Becker and Giglio (2020), and Dupraz, Nakamura and Steinsson (2021). Theoretical work includes Ilut et al. (2018), Straub and Ulbricht (2019), Kozeniauskas et al. (2018), Gilchrist and Williams (2000), Kocherlakota (2000), Hansen and Prescott (2005), Bianchi (2011), and Bianchi et al. (2017).

[8] Again, past work noted above has shown that learning can help explain stock return dynamics. The

It is also important to note that many of the statistics that the calibration matches are unconditional. The nonlinear dynamics appear all the time, not just in crashes. The paper's observation is that when investors understand that crashes can happen, that influences how they process information and thus affects the dynamics of prices even in samples in which no disasters occur.

The second part of the quantitative analysis derives nonparametric predictions from the model. First, as discussed above, the model has predictions for the relationship between volatility, its own lag, returns, and conditional skewness, that we test and find hold well in the data. Second, it is possible to estimate the precision of agents' signals and their implied uncertainty about the level of fundamentals, both without knowing the underlying model. In US stock market data, the data implies investors' conditional distribution for fundamentals has on average a standard deviation of 10.4–16.5 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 the model-based estimate (subject to the usual caveat that disagreement and uncertainty are theoretically distinct).

Implications

The paper's core claim is that understanding the wide range of nonlinearities observed in stock market returns does not require a wide range of models. The behavior is consistent with a simple setup in which the true value of stocks drifts up over time and faces occasional large negative shocks, but investors only have noisy signals about that true value. While the low-frequency behavior of the stock market is driven by deeper economic shocks, the way those movements play out at daily, weekly, and monthly time scales is consistent with what would be expected if prices reflect the evolution of investor beliefs as they progressively learn about the true state of the world.

The tools and analysis developed in this paper are applicable much more generally than just in the aggregate stock market. They naturally also have implications for the dynamics of other asset prices, but at a deep level they are really just about how beliefs evolve. The paper's methods are therefore relevant much more broadly in economics because they yield predictions for the evolution of beliefs in generic non-Gaussian settings, even in the presence of potentially severe behavioral biases. The stock market has very rich data, with both a long time series and high sampling frequency, so it is a good first setting in which to examine the relevance of this paper's type of analysis, but other economic series, such as inflation and interest rates, also display nonlinearities like skewness and correlated levels and volatility.

point here is that these results are extremely general and robust and not dependent on the specific settings studied in past work.

Additional related work

This paper is most closely related to past work studying non-Gaussian filtering problems, especially as applied to asset prices, 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 this paper's setup.[9]

While this paper is most closely related to purely Bayesian models like those above, the analysis is consistent with a number of deviations from the standard full information rational expectations setup that have been analyzed in the literature, such as misspecified or imperfect priors (e.g. Farmer, Nakamura and Steinsson (2024)). Bordalo, Gennaioli, Porta and Shleifer (2019) model diagnostic expectations as a situation in which agents distort a standard Bayesian update by treating signals as more precise than they truly are. Section 5.1 shows that this paper's analysis encompasses such models.[10] More generally, what the paper requires is that agents' update has the algebraic form of Bayes' theorem, but any of the inputs, like the prior or the data-generating processes for the latent state and signal, can be misspecified.

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).[11]

Finally, since a central driving force in this paper's analysis is the dynamic process for uncertainty, Altig, Barrero, Bloom, Davis, Meyer and Parker (2022) and Bachmann, Carstensen, Lautenbacherr, Menkhoff and Schneider (2024) are important recent 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 2 presents empirical characteristics of returns that motivate the analysis. Section 3describes the model structure and gives the

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

[10]Other biases such as excessive extrapolation can also be easily incorporated. It is also straightforward to allow for multiple priors and ambiguity aversion (e.g. Gilboa and Schmeidler (1989)) – this paper's contribution is simply to describe how the Bayesian update of each of the priors is done.

[11]See also Wachter and Zhu (2023) for a model of learning with rare disasters and Maenhout et al. (2025) for related work with ambiguity aversion. Baker, Bloom, Davis and Sammon (2025) provide evidence on what the actual events are that cause jumps in stock prices.

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.[12] All moments for realized returns are for the CRSP total market return in excess of the risk-free rate (from Kenneth French). Daily volatility is obtained by forecasting realized volatility using the VIX, so as to remove the time variation in risk premium.[13]

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 variance, skewness, and excess kurtosis of 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.

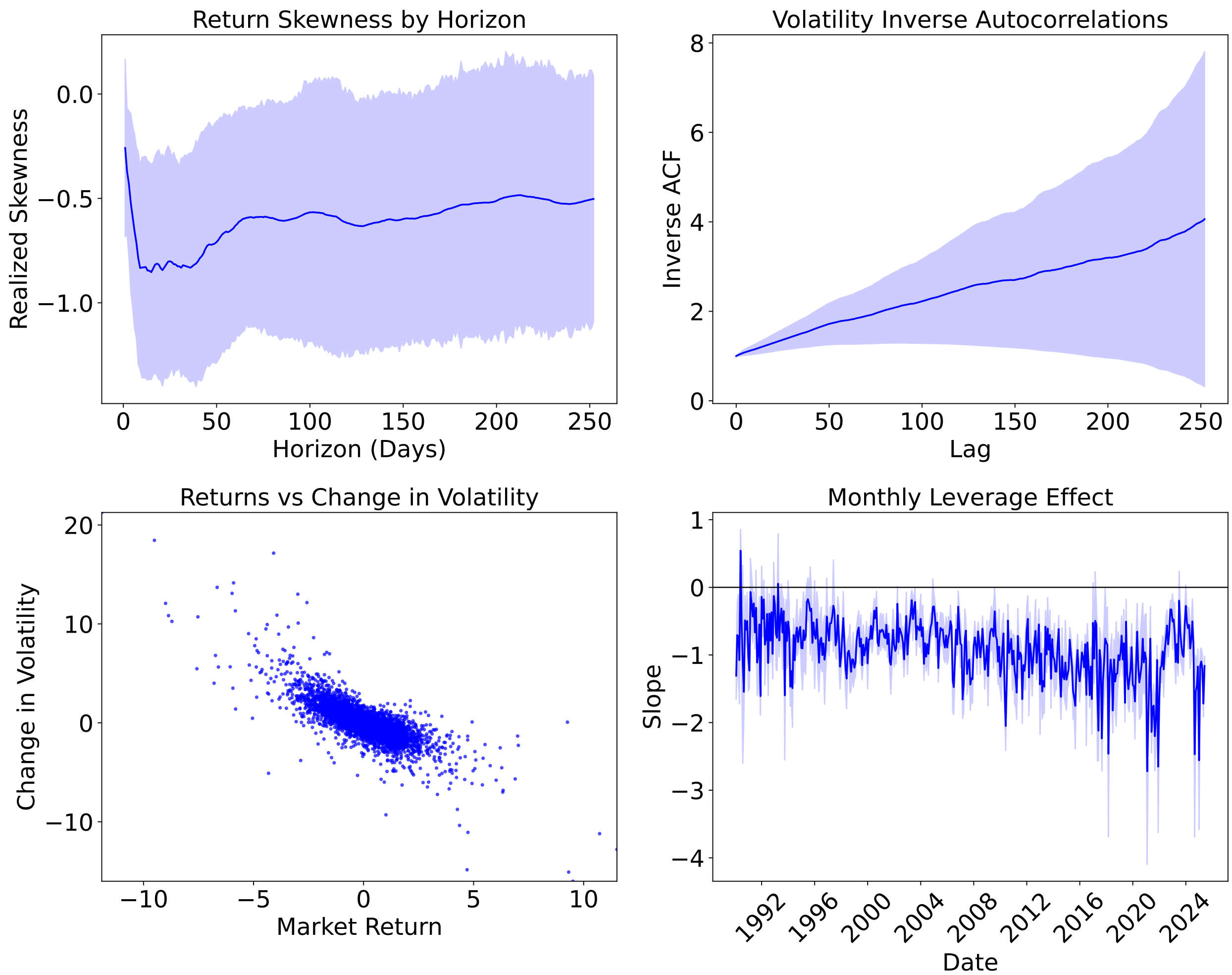
The top-left panel of figure 1 plots the skewness of realized returns over holding periods of 1 to 252 trading days. Skewness becomes significantly more negative as the horizon initially increases, and then reverts somewhat back towards zero, reaching about -0.8 for annual returns. Such a hump shape would not arise if returns were independent over time.

[12] Cont (2001) notes that similar behavior is observed across many different financial markets.

[13] The regression forecasts is conditional just on what goes into the regression, so errors will arise if the conditioning set is too small. We do not find that typical cyclical variables provide any additional forecasting power beyong the VIX.

To be more specific, the conditional volatility of returns is the projection of realized return volatility onto current option-implied volatility (the VIX index). Conditional skewness is constructed using projections of realized skewness calculated as in Neuberger (2012).

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.

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 referred to as the leverage effect. The correlation coefficient in the sample is -0.78. 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, and also has been generally trending downward.

Finally, the top-right panel of figure 1 plots the inverse autocorrelations of volatility, $1/corr\left(vol_t, vol_{t-j}\right)$. The reason to plot the inverse autocorrelations is to help visualize 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 approximately 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).[14]

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 plausible.

3 Model setup and solution

3.1 Model setup

3.1.1 Dynamics of fundamentals

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

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

where $\mathbb{E}$ is the expectation operator and 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.[15] As a simple example, a Lucas tree economy generates a pricing equation of this form. Intuitively, the simplest

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

[15] 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.

version of the analysis holds when D and M are exogenous to agents' information and we assume that for the baseline results.[16]

At any given time, the full set of information that an agent could possibly know is represented by some abstract object θ_t (which might be a scalar, a vector, a function, or something more exotic). The fundamental value of the asset is its price conditional on complete knowledge of θ_t,

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

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 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 X_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.

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. $P_t(\{Y^t, \theta_t\}) = P_t(\theta_t)$),[17] then by the law of iterated expectations,

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

(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

[16] Cases where marginal utility is a function of information either directly (e.g. generalized recursive preferences or multiplier preferences) or indirectly through a dependence of marginal utility on prices (as in the CAPM) are significantly more complicated. Section 5.3 discusses the extent to which they can fit into the paper's structure.

[17] 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.

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,

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

where x_t is the log NPV process,

$$x_t \equiv \mathbb{E}\left[\log \int_{s=0}^{\infty} \frac{D_{t+s}M_{t+s}}{M_t} ds \mid \theta_t\right] \tag{5}$$

and p_t is the **log** of the price agents actually pay, given Y^t. That act of passing the log through the expectation 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. Appendix A.3.3 reports analogs to the main results below without applying (4) and shows that they are similar, and the quantitative analysis in section 6 does not use this approximation.

Finally, for the Y process, we assume that all of the agents' information is generated via

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

where $\sigma_{Y,t}$ follows some exogenous process (subject to assumption 3 in appendix A.1.1) and the information set Y^t represents the history of Y up to date t. A reasonable benchmark is that $\sigma_{Y,t}$ is constant, but it could also vary over time, giving a form of time-varying uncertainty. When $\sigma_{Y,t} = 0$, we recover the full-information benchmark, in which $p_t = x_t$. In contrast, when $\sigma_{Y,t} > 0$, stock prices deviate from fundamentals in the short run because of the noise in investors' signals. When $\sigma_{Y,t}$ is bounded away from zero, the very short-run fluctuations in prices are dominated by learning dynamics. We maintain this assumption throughout the paper.

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.[18] The Y process can be thought of as a simplification that captures all the information agents receive in a single factor.[19] And given that x represents

[18] 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.

[19] Note that if agents receive multiple Gaussian signals, then those signals can always be combined in such a way (weighting by their precisions) that they can be reduced to a single composite signal. Normality of

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 – see section 5.2. The analysis is extremely general in the dynamics for fundamentals, represented by x, but pays for that generality with this restriction on the information structure. The paper shows that this setup can match many features of the data on the US stock market, but for other purposes, one might naturally want a different structure.

Up to equation (6), there is little loss of generality in the analysis other than the restriction that M and D are exogenous to information. The choice of the information structure is where the model is significantly restricted. Section A.4 examines how to get to equation (4) in a Lucas tree type economy with the information structure in (6).[20]

Finally, while the setup is motivated by an asset pricing problem, it is much more general. x is just some latent object of interest – it could be trend inflation, for example. Then $\mathbb{E}\left[x_t \mid Y^t\right]$ would represent agents' expectations of trend inflation given their history of signals. The following results are general statements about nonlinear filtering, not just asset price dynamics.

3.2 Solution to the filtering problem

The paper's results primarily involve the first three moments of agents' posteriors, denoted here by $\kappa_{1,t}$, $\kappa_{2,t}$, and $\kappa_{3,t}$

Proposition 1 *Given (6) and restrictions on x_t given in appendix A.1, the posterior mean ($\kappa_{1,t}$) and variance ($\kappa_{2,t}$) satisfy*

$$dp_t = d\kappa_{1,t} = \frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t\left[x_t\right]dt\right) + \mathbb{E}_t\left[dx_t\right] \tag{7}$$

$$\begin{aligned} d\,\mathrm{var}_t\left[x_t\right] &= d\kappa_{2,t} = \frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t\left[x_t\right]dt\right) - \frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt \\ &\quad + \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right] + 2\,\mathrm{cov}_t\left(x_t, dx_t\right) \end{aligned} \tag{8}$$

where $\kappa_{3,t}$ is the posterior third moment, $\langle x\rangle$ is the total quadratic variation process of x, and $\mathbb{E}_t\left[\cdot\right] \equiv \mathbb{E}\left[\cdot \mid Y^t\right]$.

the errors can be motivated by the functional central limit theorem.

[20]What makes it nonstandard is that the agent uses just the composite signal Y, and not the realization of dividends, for forming expectations. While we argue that is empirically reasonable – both because the model captures major features of the data and because empirically dividends are almost entirely predetermined – it is not completely trivial to apply to a standard setup.

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: $\kappa_{2,t}/\sigma_{Y,t}^2 = \mathrm{cov}_t(x_t, dY_t) / \mathrm{var}_t(dY_t)$ is the coefficient from a hypothetical local regression of x on dY.

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 $\kappa_{3,t} = \mathbb{E}_t\left[(x_t - \mathbb{E}_t[x_t])^3\right]$ is equal to $\mathrm{cov}_t\left((x_t - \mathbb{E}_t[x_t])^2, dY_t\right)/dt$, 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. $\mathbb{E}_t[d\langle x\rangle_t]$ represents the expected accumulation of variance in x (i.e. due to shocks), and $2\,\mathrm{cov}_t(x_t, dx_t)$ is the accumulation of uncertainty due to x effectively "spreading out" over time.

Remark 1 *Prices follow an Itô diffusion and satisfy*

$$\begin{aligned} dp_t &= \mu_{p,t}dt + \sigma_{p,t}d\tilde{W}_t \\ \text{where } d\tilde{W}_t &\equiv \sigma_{Y,t}^{-1}(dY_t - \mathbb{E}_t[x_t]\,dt) \\ \mu_{p,t}dt &= \mathbb{E}_t[dx_t],\ \sigma_{p,t} = \kappa_{2,t}/\sigma_{Y,t} \end{aligned} \tag{9}$$

and $\tilde{W}_t$ is a Brownian motion with respect to the agent's filtration.

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 $\tilde{W}$ (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. The first three cumulants are equal to the first three central moments. Denote the n-th cumulant of the time-t conditional distribution of x_t by $\kappa_{n,t}$.

Theorem 1 *Under the conditions of proposition 1, for all n for which the $n+1$th cumulant* *exists*[21]

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

where B_j denotes the jth complete exponential Bell polynomial.

The result follows from a straightforward application of textbook results in Liptser and Shiryaev (2013) and Bain and Crisan (2009), but the application to moments and cumulants is novel to this paper, as far as we can tell.[22] 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

This section examines 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 for log prices[23]

$$
\begin{aligned}
vol_t &\equiv \left(\lim_{\Delta t \downarrow 0} \mathbb{E}\left[\left(p_{t+\Delta t} - \mathbb{E}_t p_{t+\Delta t}\right)^2\right]/\Delta t\right)^{1/2} \qquad (11) \\
&= \kappa_{2,t}/\sigma_{Y,t} \qquad (12)
\end{aligned}
$$

That is, vol_t is simply the diffusive volatility of prices from the representation (9). 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.

[21]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, Wu, Shamai and Verdú (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.

[22]Theorem 1 is closely related to results in Dytso, Poor and Shitz (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 Shitz (2022) discuss, there do not appear to be any other earlier precedents to the family of results in their work and ours, but we find it unlikely that nobody else did a similar calculation at some point.

[23]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.

Combining equations (8) and (7) from proposition 1 yields the following.

Corollary 2 *Wvol_t follows a diffusion satisfying*

$$d\left(vol_t\right) = \underbrace{\sigma_{Y,t}^{-1}\frac{\kappa_{3,t}}{\kappa_{2,t}}\left(dp_t - \mathbb{E}_t dp_t\right)}_{ⓐ} - \underbrace{\frac{vol_t^2}{\sigma_{Y,t}}dt}_{ⓑ} + \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right] + 2\,\text{cov}_t\left(x_t, dx_t\right) - vol_t\sigma_{Y,t}^{-1}d\sigma_{Y,t} \quad (13)$$

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 terms on the second line again involve the spreading out of x_t itself along with the dynamics of $\sigma_{Y,t}$, both of which are exogenous, as opposed to coming from the learning that is the paper's focus.[24]

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.

4.1 The leverage effect

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

$$\frac{\text{cov}\left(dp_t, dvol_t\right)}{\text{var}\left(dp_t\right)} = \frac{\kappa_{3,t}}{\sigma_{Y,t}\kappa_{2,t}} \quad (14)$$

The term ⓐ in (13) 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

[24] In principle, $\text{cov}_t\left(x_t, dx_t\right)$ 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.

states of the world observed in our sample. The relationship has additionally strengthened over time, which would be consistent with a decline in $\kappa_{3,t}$.

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, driving down return volatility.

4.2 Slow decay in volatility

The term ⓑ in (13) 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.[25] When jumps up in vol_t are large relative to its steady-state value, its decay after a time Δt is approximately of the form $1/(1 + a\Delta t)$ for a coefficient a.[26] 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, they update less strongly and learning slows. While that phenomenon is well known in linear Gaussian filtering problems, being 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.

[25] 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.

[26] Specifically, if there is a jump at some date t_0, then if there are no further shocks to volatility and $\sigma_{Y,t}$ is constant, so that it just deterministically falls, then to leading order $y_{t_0+\Delta t} \approx y_{t_0} / \left(1 + \sigma_Y^{-1} y_{t_0} \Delta t\right)$.

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 the scaled third moment:

$$skew_t\left(p_{t+\Delta t}\right) \equiv \frac{\mathbb{E}_t\left[\left(p_{t+\Delta t} - E_t p_{t+\Delta t}\right)^3\right]}{\mathbb{E}_t\left[\left(p_{t+\Delta t} - E_t p_{t+\Delta t}\right)^2\right]^{3/2}} \tag{15}$$

Proposition 3 *The local skewness of returns is*

$$\lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right)\left(\Delta t\right)^{-1/2} = 3\frac{\kappa_{3,t}}{\kappa_{2,t}}\sigma_{Y,t}^{-1} \tag{16}$$

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, equation (16) provides a link between the conditional skewness of returns – which is potentially measurable – and the conditional skewness of fundamentals, $skew_t\left(x_t\right)$, which determines the leverage effect:

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

$$\lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right)\frac{1}{3}\left(\Delta t\right)^{-1/2} = \frac{\mathrm{cov}\left(dp_t, dvol_t\right)}{\mathrm{var}\left(dp_t\right)} \tag{17}$$

These results show 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 (13). When $\kappa_{3,t} < 0$, declines in prices raise volatility, leading to a relatively long left tail in returns. Neuberger (2012) and Neuberger and Payne (2021) study this mechanism in detail.

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 can be seen by combining equations (8) and (16) to obtain

$$std\left(vol_t\right) = \frac{1}{3}vol_t\left|\lim_{\Delta t \downarrow 0} skew_t\left(p_{t+\Delta t}\right)\left(\Delta t\right)^{-1/2}\right| + o\left(\Delta t^{1/2}\right) \tag{18}$$

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, Jonathan E. 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 Examples

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

4.5.1 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 converge to constants; 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 prices or their volatility.

4.5.2 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 proposition 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, which is 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:

$$\kappa_{2,t} = \pi_t (1 - \pi_t) \tag{19}$$

$$\kappa_{3,t} = (1 - 2\pi_t) \times \kappa_{2,t} \tag{20}$$

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 $sign\left(\kappa_{3,t}\right) = sign\left(\frac{1}{2} - \pi_t\right)$. 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 reverses: 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 effectively stops once agents know the economy is in a recession – in the sense that things cannot continue to get worse – then the leverage effect disappears or even reverses. That is not what is observed historically in the US stock market.

5 Extensions and robustness

5.1 How much rationality needs to be assumed here?

At first glance, this model appears to require investors to be strongly rational, applying Bayes theorem with full knowledge of the dynamics driving fundamentals. While that is the baseline that we will apply in the simulations below, there is nothing about the analysis that actually requires it. Proposition 1 and remark 1 hold as long as agents update using the basic form of Bayes' rule. In particular, none of the following is required in the derivations:

1. That agents use the correct precision (σ_t^2) in calculating their update (this is the deviation in the work of Bordalo, Gennaioli, Porta and Shleifer (2019) on diagnostic expectations)

2. That agents' assumed dynamics for x are in any sense "correct"

3. That the signals agents observe are truly Gaussian

4. That the agents incorporate all information they receive. They could irrationally or inefficiently ignore some information, and unreasonably privilege other sources

5. That the agents properly weight all information they receive. Agents might receive many Gaussian signals, which can be combined into a single value and used for updating. It is possible that they do that combination incorrectly

6. That agents all receive the same information or share priors. For example, their beliefs could be affected by different life histories, as in Malmendier and Nagel (2016).

What is important here is not actually that agents are true Bayesians. The propositions and corollaries above simply require that they use an updating rule that has the same algebraic structure as Bayes' rule. Obviously deviations from rational expectations will mean that what agents see as a martingale will not appear to be one to an econometrician. But the nonlinearities the paper focuses on do not hinge on anything like complete rationality.

5.2 Discrete information revelation events

Dytso, Poor and Shitz (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 moments also update by discrete amounts.

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

$$\frac{d}{dy}\kappa_j\left(x_t \mid y_t = a\right) = \kappa_{j+1}\left(x_t \mid y_t = a\right)/\sigma^2 \tag{21}$$

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 any $a \in \mathbb{R}$,

$$\mathbb{E}\left(x_t \mid y_t\right) = \sum_{j=0}^{\infty} \frac{\kappa_{j+1}\left(x_t \mid y_t = a\right)}{j!}\left(\frac{y_t - a}{\sigma^2}\right)^j \tag{22}$$

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 – continuity is not essential, but proposition 4 still requires a normally distributed signal.[27] That said, proposition 4 also shows why continuous time is useful here: in (21) the posterior cumulants, which are precisely what one wants to solve for, determine sensitivity. So using (21) really requires solving for a fixed point. In continuous time the cumulants follow continuous processes, so the prior and posterior values are effectively identical, which simplifies the analysis. Additionally, though, note that theorem 1 does not follow directly

[27] 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.

from proposition 4. Accounting for dynamics, which is ultimately central to the analysis, makes the derivations significantly more complicated.

5.3 Allowing marginal utility or cash flows to depend on information

The main results restrict to the case where state prices are not affected by the realization of the signal that agents observe. In some of the learning models in the literature, there is an additional mechanism that enters because information is priced (i.e. M depends on the signal itself). This paper rules that channel out because it adds significant complications to the analysis and the goal is to focus on the implications that follow directly the learning, rather than additional price effects.

In particluar, there are two ways that the assumption that information is unpriced can be relaxed. The first is to allow marginal utility and cash-flows, M and D, to depend on the signal errors W in a known way. Then the fundamental value X is correlated with W, which is allowed by the standard filtering results that we use from Liptser and Shiryaev (2013). There is just a simple a correction to the gain coefficient.

In models in which information is endogenously priced, though, such as under the CAPM or generalized recursive preferences, that price is not a known exogenous function. Rather, finding it involves solving a fixed point problem. In the CAPM for example, the response of prices to a shock depends partly on how future volatility responds. But the response of future volatility itself depends on how prices respond to shocks.[28] While that class of models is important in the literature, the fixed point adds significant complexity (enough that the model is no longer solvable by hand)[29] and the endogeneity is not necessary for matching the features of the data that motivate the analysis. There is nothing about that fixed point structure, though, that appears impossible in principle to incorporate into this paper's type of analysis in future work. See appendix A.4.2 for a discussion.

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

[28] The information structure from Veronesi (1999) that is studied in section 4.5.2 is a special case of what is studied here, but the full model in Veronesi (1999) involves exactly this fixed point, which in the end is solved numerically.

[29] E.g. see Veronesi (1999).

qualitative predictions above map into quantitatively reasonable behavior. The example shows how layering incomplete information over simple dynamics for fundamentals can generate quantitatively 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 analyzed here that would do better, just that the exact specification detailed in this section is (obviously) imperfect.

6.1 Model setup

Fundamentals have an average growth rate of g with both small Gaussian shocks and occasional downward jumps:

$$dx_t = (\phi\lambda + g)\, dt - J_t dN_t + \sigma_x dB_t \tag{23}$$

where B is a Brownian motion, N is a Poisson process with constant rate ϕ, and J_t is a random variable with mean λ. The term $\phi\lambda dt$ ensures that mean price growth is equal to g.[30] The drift g can be thought of (and formally motivated as) coming from a risk premium on cash flows. It plays no role other than to generate positive average returns.

In the absense of learning, prices are equal to x and hence inherit its dynamics. So without learning, returns would be skewed due to the jumps, but volatility would be constant and there would be no leverage effect or long memory.

The free parameters are those determining the distribution of jumps, the diffusive volatility, and the error variance of the signals (σ_Y^2). For the jumps, we use the calibration of Barro and Jin (2011) and assume that equities have a leverage of 3 relative to shocks to consumption.[31] σ_x is then chosen to match the standard deviation of US stock returns, and $\sigma_Y = 5.5$ (based on the time index corresponding to days) was chosen relatively roughly to match the dynamics of volatility and skewness. 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 0.35 in logs. We set

[30] 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).

[31] This is high relative to the average debt-to-equity ratio for US stocks. On the other hand, large negative shocks mechanically raise leverage, so a value of 3 can be seen as representing more how leverage would act in a disaster state. This value is consistent with that chosen by Bansal and Yaron (2004).

While Barro and Jin (2011) cut their jump distribution off at a minimum of 10.5 percent for consumption, we find that extending it to zero helps the model fit the data. Barro and Jin (2011) do not rule out the existence of small jumps but rather just focus on the larger ones.

$g = 0.07$ to match the historical equity premium.

It is straightforward to simulate the model by discretizing the state space and then calculating expectations using Bayes' theorem.[32] In addition, prices are calculated as $P_t = \mathbb{E}_t \exp(x_t)$, so that the simulation does not use the log approximation from equation (4).

6.2 Impulse response functions

This section examines two impulse response functions – to errors in the signal, $\sigma_Y dW_t$, and to jumps in fundamentals, $J_t dN_t$. Since the model is nonlinear, impulse responses differ depending on the state of the economy. Impulse responses here are calculated simply as the population mean (from a 100,000-year simulation) conditional on a shock having occurred compared to the population mean conditional on a shock not having occurred.

The noise shock is defined as a month with a ± 2.4 standard deviation realization in the total error in the signal over the month, representing a one-in-ten year event.[33] 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. First, consider the negative shock. As agents observe the negative signals, prices fall and uncertainty rises. 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 because when uncertainty is low, agents update less in response to the incoming data that is (by assumption) telling them that a disaster did not actually happen.

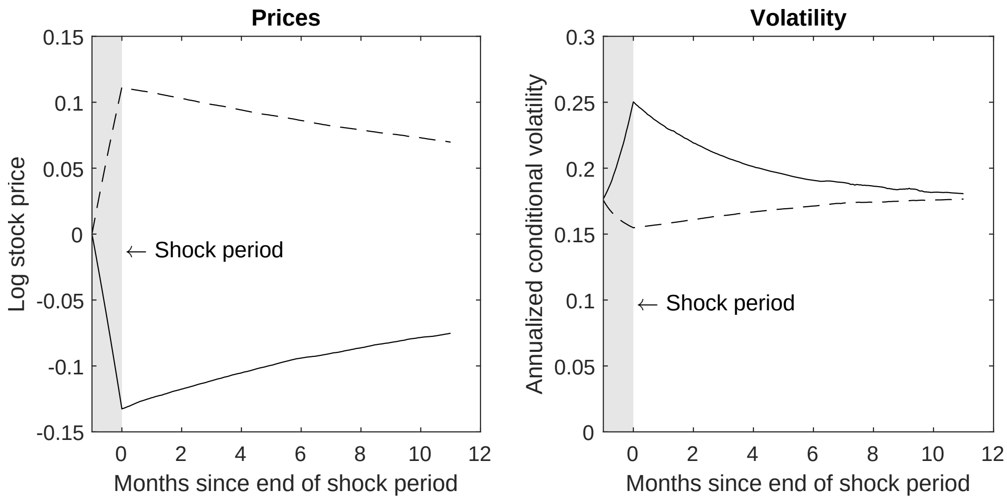
In the case of a positive noise shock, the effects are smaller. Intuitively, that happens because of the asymmetry in the fundamentals process: agents understand that it is possible that a large negative shock has occurred, so they update strongly in response to negative signals. But since large positive shocks are effectively impossible, they strongly discount positive signals. That plays out by volatility declining, so that the marginal effects of additional noise are progressively smaller. The overall effect on prices of a negative noise

[32] 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. Specifically, the grid for the state space has increments of one log point and the time interval is set to a day.

[33] More specifically, the IRFs are calculated as average behavior following a month in which the cumulative error was between ± 2.35 and ± 2.45 standard deviations.

shock is 20% larger than a positive noise shock, and the effect on volatility is 3.5 times larger.

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 $\sigma_Y dW$).

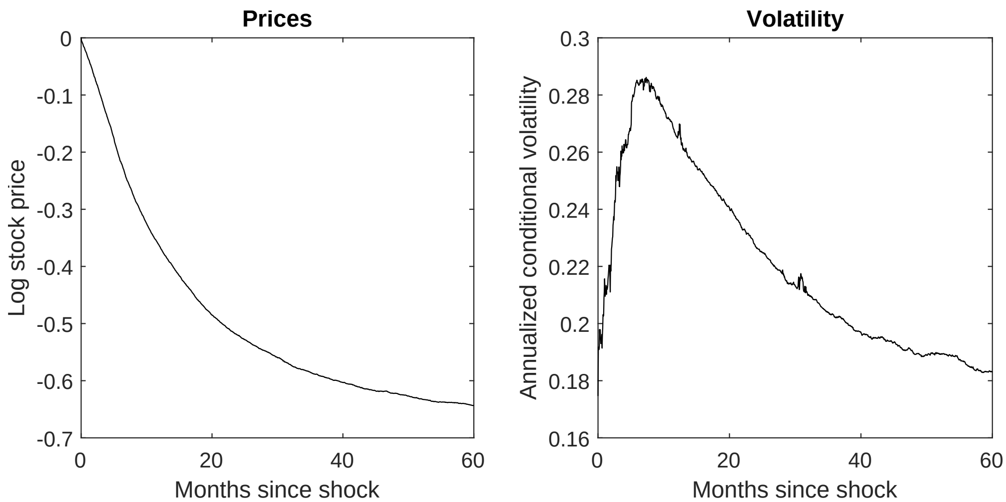
The third IRF, in figure 3, represents the average response to a downward jump in fundamentals. The left-hand plot in figure 3 shows that the decline in prices is again nonlinear – it accelerates before slowing, with the initial declines on average equal to 0.15 percent per day, accelerating to 0.20 percent per day at their peak. It takes at least five years on average for prices to fully incorporate the drop in fundamentals. 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.

The right-hand panel shows that the peak in volatility does not come on average until 7-8 months after the shock. That said, because of the nonlinearity of the model, the mean IRF is not a very good representation of typical behavior. For example, the average maximum of annualized volatility in the five years after a negative jump is 72 percent, more than twice as high as the average five-year peak of only 35 percent for periods with no jump.

6.3 Simulation results

The left-hand panel of figure 4 plots an example of a 100-year time-series of fundamentals, x, and prices, p, from the full simulation. Prices track fundamentals well in the long-run, but clearly there can be large temporary deviations, and those deviations are skewed left. In some cases, fundamentals jump down and it takes time for prices to catch up, and there

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.

are also clear cases of prices jumping down "erroneously" (based on hindsight or on knowing the true state) and then recovering, for example around year 50.

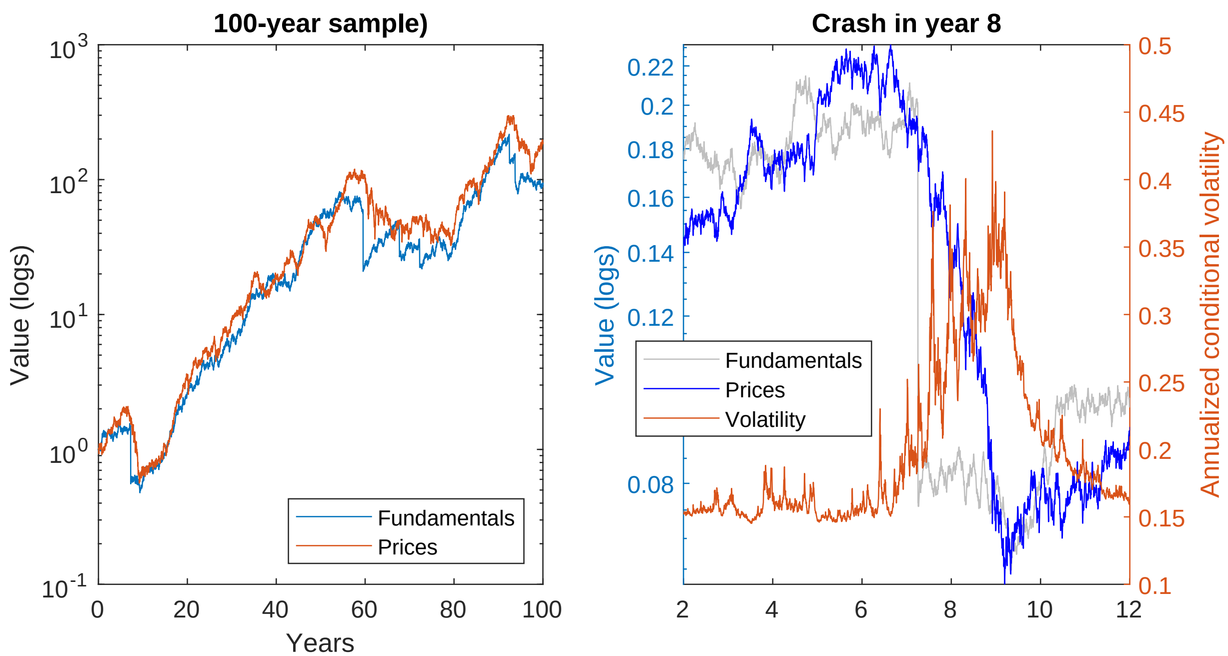
The right-hand panel of figure 4 plots the dynamics of prices and conditional volatility during the crash observed around year 8 in the simulation in the left-hand panel. The crash displays what can be recognized as common behavior in the data, with the price decline accelerating and then ending with a large rebound from the bottom. Volatility rises as prices fall, peaking at the bottom, and then falling as prices recover somewhat.

Table 1 reports moments for returns and their conditional volatility. As discussed in Barro and Jin (2011), the US has historically had fewer disasters than would be expected unconditionally. We therefore calculate moments as the average from 100-year samples of the simulation in which the most negative annual return is no more negative than the most negative value observed in the US over the past century. That choice leads to somewhat less skewness in the simulations. Calculating moments on 100-year samples also accounts for any small-sample bias in the statistics.

Table 1 shows that model broadly matches the data, though skewness and kurtosis for volatility are somewhat higher than their empirical counterparts. Since the goal of this section is just to present a parsimonious quantitative example, the good match is fairly surprising.

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

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.

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 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.

Table 2: Model vs data moments

	Stock returns		Volatility level		Volatility change	
Moment	Data	Model	Data	Model	Data	Model
Std. dev.	1.14	1.11	6.64	6.74	1.39	0.93
Skewness	-0.26	-0.10	2.19	5.11	1.51	1.03
Kurtosis	12.68	4.75	11.57	43.52	30.03	139.05
Corr. w/ R_t					-0.78	-0.67

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

7.1 Regression setup

Combining equations (12), (8), and (16), and assuming the price is a martingale and holding $\sigma_{Y,t}$ constant, we have

$$d\left(vol_t\right) = \left(\frac{1}{3}\Delta t^{-1/2}\right) skew_t\left(p_{t+\Delta t}\right) dp_t - \frac{1}{\sigma_{Y,t}} vol_t^2 dt + \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right] \tag{24}$$

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

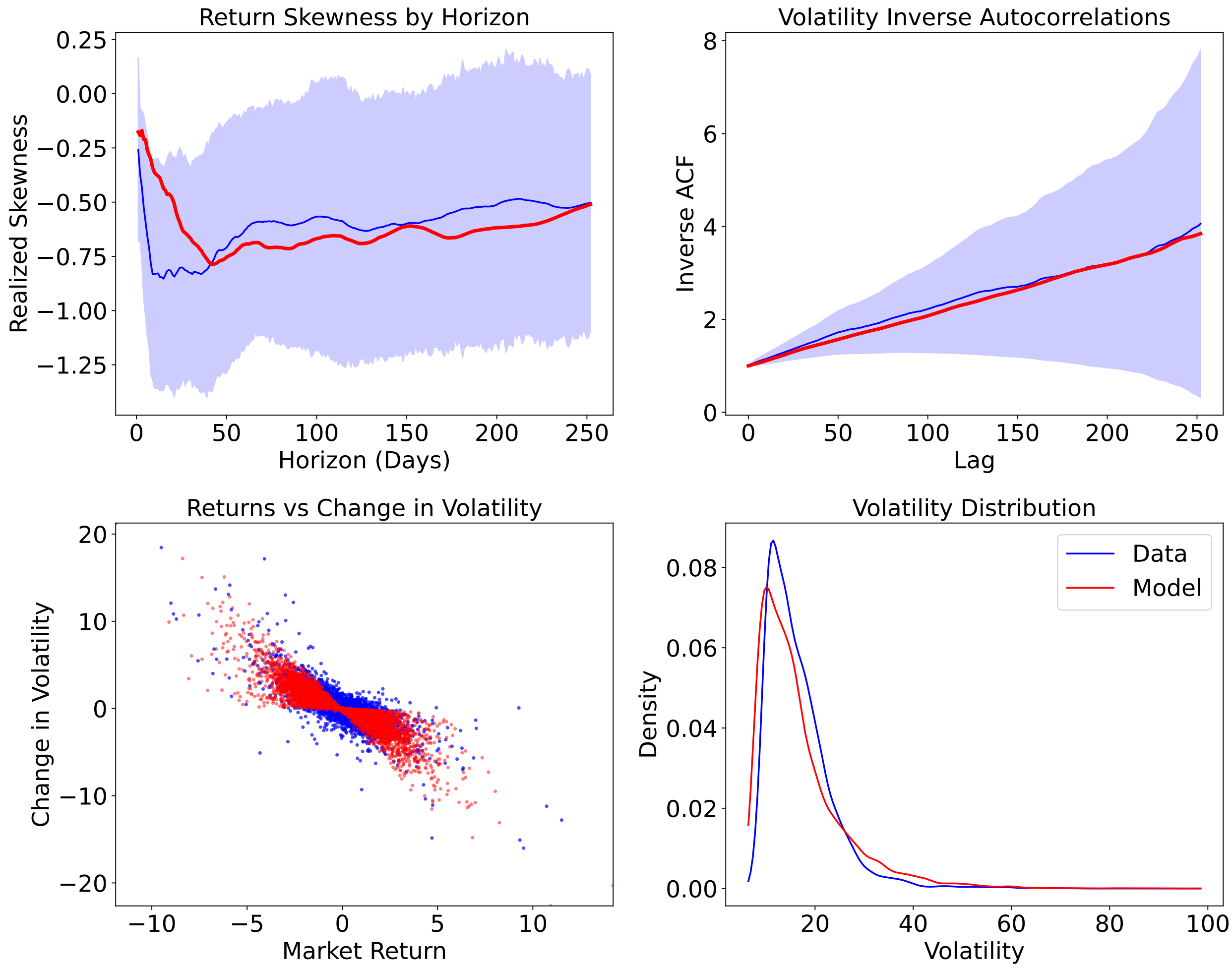
The model has two predictions for the results of this regression: the coefficient on $\left(\frac{1}{3}\Delta t^{-1/2}\right) skew\left(p_{t+\Delta t}\right) dp_t$ should be equal to 1, and the coefficient on $vol_t^2 dt$ is equal to $\sigma_{Y,t}^{-1}$. The first represents a test of the model. That relationship in fact holds as long as prices and volatility follow a joint diffusion driven by a single Brownian motion, and so it tests that aspect of the model. The second prediction shows that the regression can be used to identify one of the model's structural parameters.

7.2 Data

We estimate the regression (24) for two markets. The first, given our focus on the stock market, is the S&P 500. For that case, we proxy for vol_t 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.

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, in contrast, consistently positive skewness. To calculate dp_t in this

Figure 5: Simulation results



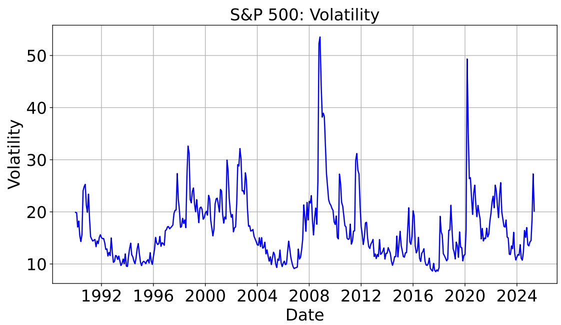
Note: The top left, top right and bottom left panels are the same as in figure 1, overlayed with the corresponding outputs from a 100-year simulation of the model based on the parameters discussed in the text. The bottom-right panel shows the estimated density of volatility in the data and in the model.

case we use natural gas futures returns from the CME. We then use options on futures to estimate the conditional moments in the same way as for the S&P 500.[34] 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.

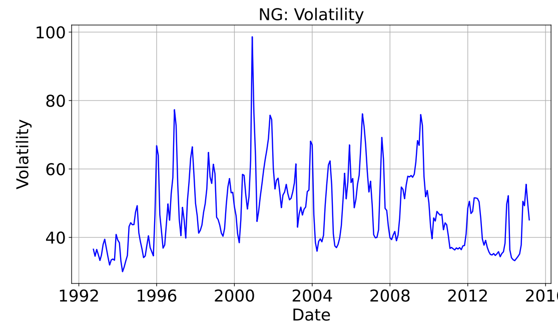
In addition to the differences in skewness, the S&P 500 is generally viewed as being significantly more important systematically – since it represents a nontrivial part of aggregate wealth – so if one was concerned about the results for the S&P 500 being somehow contaminated

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

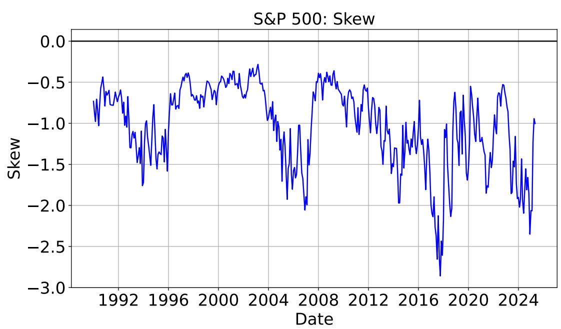
Figure 6: Time series of VIX and Skew for S&P 500 and natural gas



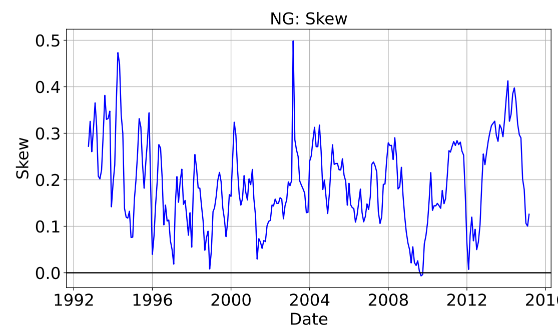
(a) Volatility (S&P 500)



(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 averages of the corresponding daily series.

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 in at least two notable cases they map somewhat reasonably into the data.

7.3 Results

Table 3 reports results of the regression implied by (24). The first and third columns report the baseline results. The coefficients are highly statistically significant and have the expected signs.

Under the model, if the various assumptions made to derive the regression setup 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 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.

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_{t-1}^2	-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.

To evaluate the relevance of the model-implied nonlinearity 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} indicating that vol_{t-1}^2 has greater explanatory power than vol_{t-1}. For the S&P the t-statistic for vol_{t-1}^2 is larger than that for $vol_{t-1}$by a factor of more than 10, but it is only marginally larger for natural gas.

7.4 Estimates of investors' uncertainty about fundamentals

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

$$vol_t = \frac{\kappa_{2,t}}{\sigma_{Y,t}} \Delta t^{1/2} \tag{25}$$

$$\Rightarrow \kappa_{2,t}^{1/2} = \left(vol_t \sigma_{Y,t} \Delta t^{-1/2} \right)^{1/2} \tag{26}$$

and that the scaling of the estimates is for a unit time interval being equal to a day.

$\sigma_{Y,t}$ is between 0.94 and 2.36, based on the coefficient on vol_{t-1}^2 in equation (24). The first thing to note is that that value is lower than the value that works well in the calibration studied above, showing that the calibration is at least somewhat misspecified for volatility dynamics.

Taking equation (26) and inserting the value for σ_Y along with the the historical daily standard deviation of stock returns in our sample, 1.05 percent, 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 (16) 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.

These estimates are both notable because they are independent of the model for x – in that sense they are model free. What they depend on is just the information structure the paper assumes, which is that prices are driven by a single composite signal that is Gaussian conditional on x.

7.5 Comparing to survey data

We have not found a survey that directly measures investors' uncertainty about fundamentals and would allow us to validate the estimate of σ_Y – i.e. a survey that asks 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.

7.6 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 model's prediction of a coefficient of 1 on the correctly scaled interaction of returns with conditional skewness also fits well, showing that conditional skewness controls the magnitude of the leverage effect over time. Finally, the coefficients themselves can be mapped into an estimate of σ_Y, the noise in investors' signals.

8 Conclusion

This paper's main results are fundamentally about how information affects beliefs in a very simple but standard Bayesian filtering setting. The analysis is motivated by the highly nonlinear behavior of the stock market, and it shows that belief dynamics under information acquisition explain many of those nonlinearities, both qualitatively and quantitatively.

The results are much more broadly applicable, though. Obviously there are many other financial markets that display different forms of nonlinearity and it is natural to ask how well the filtering mechanism works in those settings. Discrete information revelation events play no role in the analysis of this paper, but are certainly much more important for individual stocks.

But information acquisition problems are pervasive in economics, the assumptions of linearity and Gaussianity are not always reasonable approximations to the data. Inflation, for example, has historically been highly skewed, with its volatility being positively correlated with its level. The analysis here may therefore be useful in understanding the evolution of inflation expectations in the face of that nonlinearity.

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.

Abreu, Dilip and Markus K Brunnermeier, "Bubbles and crashes," *Econometrica*, 2003, *71* (1), 173–204.

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.

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.

Bansal, Ravi and Amir Yaron, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 2004, *59* (4), 1481–1509.

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

Barro, Robert J and Tao Jin, "On the size distribution of macroeconomic disasters," *Econometrica*, 2011, *79* (5), 1567–1589.

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.

Bloom, Nicholas, "The Impact of Uncertainty Shocks," *Econometrica*, 2009, *77(3)*, 623–685.

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.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, "Diagnostic expectations and stock returns," *The Journal of Finance*, 2019, *74* (6), 2839–2874.

Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek, "The macroeconomic impact of financial and uncertainty shocks," *European Economic Review*, 2016.

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.

Cont, Rama, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative finance*, 2001, *1* (2), 223.

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," 2021. 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.

Gilboa, Itzhak and David Schmeidler, "Maxmin Expected Utility with Non-unique Prior," *Journal Of Mathematical Economics*, 1989, *18* (2), 141–153.

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

Gormsen, Niels Joachim and Christian Skov Jensen, "Higher-Moment Risk," 2023. Working paper.

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.

Iseringhausen, Martin, Ivan Petrella, and Konstantinos Theodoridis, "Aggregate skewness and the business cycle," 2023, pp. 1–38.

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.

Liptser, Robert S and Albert N Shiryaev, *Statistics of random processes: I. General theory*, Springer Science & Business Media, 2013.

Ludvigson, Sydney C, Sai Ma, and Serena Ng, "Uncertainty and business cycles: exogenous impulse or endogenous response?," *American Economic Journal: Macroeconomics*, 2021, *13* (4), 369–410.

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

Maenhout, Pascal J., Hao Xing, and Anne G. Balter, "Model Ambiguity versus Model Misspecification in Dynamic Portfolio Choice," 2025. Forthcoming, Journal of Finance.

Malmendier, Ulrike and Stefan Nagel, "Learning from inflation experiences," *The Quarterly Journal of Economics*, 2016, *131* (1), 53–87.

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.

Menkhoff, Manuel, "The Devil Is in the Tail: Macroeconomic Tail Risk Expectations of Firms," 2025. 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.

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.

Phillips, Peter CB, Shuping Shi, and Jun Yu, "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International economic review*, 2015, *56* (4), 1043–1078.

Protter, Philip E., *Stochastic Integration and Differential Equations*, Springer, 2005.

Reinhart, Carmen M. and Kenneth S. Rogoff, *This time is different: Eight centuries of financial folly*, Princeton University Press, 2009.

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

Salgado, Sergio, Fatih Guvenen, and Nicholas Bloom, "Skewed Business Cycles," 2020. Working paper.

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," 2019, *183*, 625–660.

Sufi, Amir and Alan M Taylor, "Financial crises: A survey," *Handbook of international economics*, 2022, *6*, 291–340.

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 Proof of theorem 1

A.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.1.2 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*

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

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

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

where

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

Or equivalently,

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

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, $\mathbb{E}_t[\mathcal{G}_{\omega,t}]dt = \mathbb{E}_t[d\exp(i\omega x_t)]$.

[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.

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 $k \leq n$,*

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

where B_j denotes the jth complete exponential Bell polynomial.

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

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

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] Taking the Itô differential of the above equation (and applying the dominated convergence theorem to switch the order of d and $d^k/d\omega^k$),

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

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

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

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

$$\frac{d^k}{d\omega^k}\frac{\mathbb{E}_t[d\exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} = \sum_{j=0}^{k}\binom{k}{j}\left(\frac{d^{k-j}}{d\omega^{k-j}}\mathbb{E}_t[d\exp(i\omega x_t)]\bigg|_{\omega=0}\right)\left(\frac{d^j}{d\omega^j}\left(\mathbb{E}_t[\exp(i\omega x_t)]\right)^{-1}\bigg|_{\omega=0}\right). \tag{A.16}$$

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 can be used to transform the right-hand side of (A.16) into (see Comtet (1974) section 3.3)

$$\begin{aligned}\frac{d^k}{d\omega^k}\frac{\mathbb{E}_t[d\exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} &= i^k\sum_{j=0}^{k}\binom{k}{j}B_j\left(-\kappa_{1,t},\dots,-\kappa_{j,t}\right)\mathbb{E}_t[d(x_t^{k-j})] &\text{(A.17)}\\ &= i^k\sum_{j=1}^{k}\binom{k}{j}B_{k-j}\left(-\kappa_{1,t},\dots,-\kappa_{k-j,t}\right)\mathbb{E}_t[d(x_t^{j})]. &\text{(A.18)}\end{aligned}$$

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

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

Therefore,

$$\frac{d^k}{d\omega^k}\frac{\mathrm{cov}_t(x_t,\exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\bigg|_{\omega=0} = i^{-1}\frac{d^{k+1}}{d\omega^{k+1}}\log\varphi_{x,t}(\omega)\bigg|_{\omega=0} = i^k\kappa_{k+1,t}, \tag{A.20}$$

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

$$i^k\frac{\kappa_{k+1,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t[x_t]dt\right). \tag{A.21}$$

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

$$
\begin{aligned}
&\frac{d^k}{d\omega^k}\left(\frac{\operatorname{cov}_t(x_t,\exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\right)^2\Bigg|_{\omega=0} && \text{(A.22)}\\
&\quad=\sum_{j=1}^{k-1}\binom{k}{j}\left(\frac{d^j}{d\omega^j}\frac{\operatorname{cov}_t(x_t,\exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\Bigg|_{\omega=0}\right)\left(\frac{d^{k-j}}{d\omega^{k-j}}\frac{\operatorname{cov}_t(x_t,\exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]}\Bigg|_{\omega=0}\right) && \text{(A.23)}\\
&\quad=\sum_{j=1}^{k-1}\binom{k}{j} i^j\kappa_{j+1,t} i^{k-j}\kappa_{k-j+1,t} && \text{(A.24)}\\
&\quad=i^k\sum_{j=2}^{k}\binom{k}{j-1}\kappa_{j,t}\kappa_{k-j+2,t}. && \text{(A.25)}
\end{aligned}
$$

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

A.2 Proof of proposition 1

The dynamics of the first moment are exactly as in theorem 1. For the second moment, we simply need to show that

$$\sum_{j=1}^{2}\binom{2}{j} B_{2-j}\left(-\kappa_{1,t},\ldots,-\kappa_{2-j,t}\right)\mathbb{E}_t[d(x_t^j)] = \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right] + 2\operatorname{cov}_t\left(x_t, dx_t\right). \tag{A.26}$$

The left-hand side of the above expression is given by

$$\mathbb{E}_t[d(x_t^2)] - 2\mathbb{E}_t[x_t]\mathbb{E}_t[dx_t]. \tag{A.27}$$

Assumption 1 implies that x_t is a semimartingale of the form $x_t = x_0 + M_t + A_t$, where M_t is a right-continuous martingale and A_t is an absolutely continuous process. By Itô's lemma for semimartingales (e.g., Theorem 32 of Protter (2005)),

$$\mathbb{E}_t[d(x_t^2)] = 2\mathbb{E}_t[x_{t^-}dx_t] + \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right], \tag{A.28}$$

where $x_{t^-} \equiv \lim_{s\uparrow t} x_s$. By assumption 1, $x_t = x_{t^-} + \Delta M_t$, where ΔM_t denotes the time-t jump of M. Since Y_t is a continuous process, its natural filtration is continuous. Therefore,

$$\mathbb{E}_t[\Delta M_t] = \mathbb{E}_t[\Delta M_t|Y^t] = \mathbb{E}_t[\Delta M_t|Y^{t^-}] = 0, \tag{A.29}$$

where the last equality is a consequence of the fact that M is a martingale. Thus, $\mathbb{E}_t[x_t] = \mathbb{E}_t[x_{t-}]$, and so

$$\mathbb{E}_t[d(x_t^2)] - 2\mathbb{E}_t[x_t]\mathbb{E}_t[dx_t] = \mathbb{E}_t\left[d\left\langle x\right\rangle_t\right] + 2\,\mathrm{cov}_t\left(x_t, dx_t\right), \tag{A.30}$$

where $\mathrm{cov}_t\left(x_t, dx_t\right) \equiv \mathbb{E}_t[x_{t-}dx_t] - \mathbb{E}_t[x_{t-}]\mathbb{E}_t[dx_t]$.

A.3 Additional derivations

A.3.1 Equation (16)

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

$$\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2] = \kappa_{2,t}^2\sigma_{Y,t}^{-2}\Delta t + O(\Delta t^2), \tag{A.31}$$

and

$$\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3] = \mathbb{E}_t\left[\left(\kappa_{2,t}\sigma_{Y,t}^{-1}\Delta W_t + \frac{\kappa_{3,t}\sigma_{Y,t}^{-1}}{2}(\Delta W_t^2 - \Delta t)\right)^3\right] + O(\Delta t^{5/2}) \tag{A.32}$$

$$= 3\kappa_{2,t}^2\kappa_{3,t}\sigma_{Y,t}^{-4}\Delta t^2 + O(\Delta t^{5/2}). \tag{A.33}$$

Therefore,

$$skew_t\left(p_{t+\Delta t}\right)(\Delta t)^{-1/2} = \frac{\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3]\,(\Delta t)^{-1/2}}{\left(\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2]\right)^{3/2}} \tag{A.34}$$

$$= \frac{\kappa_{3,t}}{\kappa_{2,t}}\sigma_{Y,t}^{-1} + O(\Delta t^{1/2}). \tag{A.35}$$

A.3.2 Equation (24)

Starting from equation (8),

$$d\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right) = \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t\left[x_t\right]dt\right) + \mathbb{E}_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.36}$$

$$= \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\kappa_{2,t}}\frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t\left[x_t\right]dt\right) + \mathbb{E}_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.37}$$

$$= \frac{1}{\sigma_{Y,t}^2}\left(skew_{t\to t+\Delta t}\left(dp_t\right)\left(\Delta t\right)^{-1/2}\right)\frac{1}{3}\frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - \mathbb{E}_t\left[x_t\right]dt\right) + \mathbb{E}_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt \tag{A.38}$$

$$= \frac{1}{\sigma_{Y,t}^2}\left(skew_{t\to t+\Delta t}\left(dp_t\right)\left(\Delta t\right)^{-1/2}\right)\frac{1}{3}\left[dp_t - \mathbb{E}_t dx_t\right] + \mathbb{E}_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}}{\sigma_{Y,t}}\right)^2 dt, \tag{A.39}$$

where the third line uses equation (16) and the fourth line inserts the formula for $dp_t = d\kappa_{1,t}$.

A.3.3 Solving the filtering problem in logs

This section analyzes the filtering problem without applying the approximation in equation (4) (and we note here again that the simulation in section 6 does not use equation (4) either). Specifically, instead of assuming $p_t = \mathbb{E}_t\left[x_t\right]$, we set $p_t \equiv \log \mathbb{E}_t\left[\exp(x_t)\right]$. Then we obtain the following counterpart to corollary 2.

Proposition 5 *When $\sigma_{Y,t}$ is constant, vol_t follows a diffusion satisfying*

$$\begin{aligned} d\left(vol_t\right) = {} & \frac{1}{\sigma_{Y,t}}\sum_{k=2}^{\infty}\frac{1}{(k-1)!}\sum_{j=1}^{k}\binom{k}{j}B_{k-j}\left(-\kappa_{1,t},\ldots,-\kappa_{k-j,t}\right)\mathbb{E}_t[d(x_t^j)] \\ & + \frac{1}{\sigma_{Y,t}^3}\sum_{k=2}^{\infty}\frac{\kappa_{k+1,t}}{(k-1)!}\left(dY_t - \mathbb{E}_t[x_t]dt\right) \\ & - \frac{1}{2\sigma_{Y,t}^3}\sum_{k=2}^{\infty}\frac{1}{(k-1)!}\sum_{j=2}^{k}\binom{k}{j-1}\kappa_{j,t}\kappa_{k-j+2,t}dt. \end{aligned} \tag{A.40}$$

In this case, instead of just depending on the third moment, the sign of the leverage effect is more complicated. In particular, the analog to proposition 2 is

$$\frac{\operatorname{cov}\left(dp_t, dvol_t\right)}{\operatorname{var}\left(dp_t\right)} = \frac{\sum_{k=2}^{\infty}\frac{\kappa_{k+1,t}}{(k-1)!}}{\sum_{k=1}^{\infty}\frac{\kappa_{k+1,t}}{k!}}\sigma_{Y,t}^{-1} \tag{A.41}$$

and the equation in proposition 3 then becomes

$$\lim_{\Delta t \downarrow 0} skew_t \left(p_{t+\Delta t}\right) \left(\Delta t\right)^{-1/2} = 3 \frac{\sum_{k=2}^{\infty} \frac{\kappa_{k+1,t}}{(k-1)!}}{\sum_{k=1}^{\infty} \frac{\kappa_{k+1,t}}{k!}} \sigma_{Y,t}^{-1} \tag{A.42}$$

What is in the main text has only the leading terms from the two power series. Note also that these together immediately imply that corollary 3 holds without any changes.

In moving on to the empirical analysis, we have the following modification.

Corollary 5 *The analog to equation (24) is*

$$\begin{aligned} d\left(vol_t\right) = &\left(\frac{1}{3}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right) \left(dp_t - \mathbb{E}_t[dp_t]\right) \\ &- vol_t^2 \left(\frac{\kappa_{2,t}}{\sigma_{Y,t}^2 vol_t} + \left(\frac{1}{3}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right)\right) dt \\ &+ \frac{1}{\sigma_{Y,t}} \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \sum_{j=1}^{k} \binom{k}{j} B_{k-j}\left(-\kappa_{1,t}, \ldots, -\kappa_{k-j,t}\right) \mathbb{E}_t[d(x_t^j)]. \end{aligned} \tag{A.43}$$

The first term is identical to the first term in regression (24). In other words, the $p_t = \mathbb{E}\left[x_t \mid Y^t\right]$ approximation has no bearing for the relationship between the leverage effect and return skewness. The second and third terms are different than those in (24). However, assuming that $\kappa_{k,t}$ is small for all $k \geq 4$, the price is a martingale, and x_t has stationary and independent increments, we have

$$\begin{aligned} d\left(vol_t\right) \approx &\left(\frac{1}{3}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right) dp_t - \frac{1}{\sigma_{Y,t}} \left(1 + \sigma_{Y,t} \left(\frac{1}{6}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right)\right) vol_t^2 dt \\ &+ \text{constant} \cdot dt, \end{aligned} \tag{A.44}$$

which is identical to (24), except for the $\sigma_{Y,t}\left(\frac{1}{6}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right)$ correction in the coefficient of vol_t^2. The $p_t = \mathbb{E}_t\left[x_t\right]$ approximation introduces a Jensen's inequality error term that has an expansion in the cumulants; $\sigma_{Y,t}\left(\frac{1}{6}\Delta t^{-1/2}\right) skew_t \left(p_{t+\Delta t}\right)$ is the leading term in that expansion.

A.4 Accommodating the information structure in an equilibrium model

This section has two parts. First, it shows how to derive equations (4) and (5) in a Lucas tree economy in which agents price assets based on a single signal, Y. The basic setup in equation (1) holds in such models, but the restriction to a single signal in the completely general case is less obvious. Second, it gives a description of how the information structure assumed in section 3 can be incorporated into more general models.

A.4.1 Lucas tree economy

There is a single tree with a cash-flow of D_t. Agents are all identical. They are endowed with a unit claim on the tree, which pays D_t in each period. Their date-t budget constraint is

$$\mathbb{E}_t \int_{j=0}^{\infty} M_{t+j} C_{t+j} dj = M_t P_t + \mathbb{E}_t \int_{j=0}^{\infty} M_{t+j} D_{t+j} dj \tag{A.45}$$

where M_{t+j} is the price of a date-$t+j$ Arrow–Debreu security.

The agents' objective is

$$\max \mathbb{E}_t \left[\int_{j=0}^{\infty} \beta^j u \left(C_{t+j} \right) dj \mid Y^t \right] \tag{A.46}$$

where u represents utility over consumption, C, and Y^t is the history of the signal process up to date t. We assume that agents' trading decision on date t, as represented by their holdings of claims on the tree, must be measurable with respect to Y^t. Their consumption is then a residual.

Consider a perturbation at the optimum that purchases one additional (infinitesimal) unit of the tree on date t – raising consumption in proportion to D_{t+j} on all future dates, and reducing consumption in proportion to P_t on date t. At the optimum, it must be the case that

$$\mathbb{E}\left[P_t u' \left(C_t \right) \mid Y^t \right] = \mathbb{E}\left[\int_{j=0}^{\infty} u' \left(C_{t+j} \right) D_{t+j} dj \mid Y^t \right] \tag{A.47}$$

The interpretation of the right-hand side is standard. The left-hand side is more subtle. It says that the cost of the purchase of an additional unit of the tree is equal to the expected marginal utility of consumption conditional on the value of Y^t, since it will reduce consumption by P_t in all states of the world in which Y^t takes on that particular value.

Rearranging and noting that P_t can only be a function of Y^t,

$$P_t = \frac{\mathbb{E}\left[\int_{j=0}^{\infty} u'\left(C_{t+j}\right) D_{t+j} dj \mid Y^t\right]}{\mathbb{E}\left[u'\left(C_t\right) \mid Y^t\right]} \tag{A.48}$$

As a first observation note that if cash-flows are pre-determined and thus measurable with respect to Y^t, then $\mathbb{E}\left[u'\left(C_t\right) \mid Y^t\right] = u'\left(C_t\right)$ and equation (A.48) reduces to equation (1) with $M_t \equiv u'\left(C_t\right)$ and hence (4) and (5) follow immediately.

Alternatively, suppose cash-flows are not predetermined. Then consider equation (5) with $M_t \equiv u'\left(C_t\right)$:

$$\begin{aligned} x_t &= \mathbb{E}\left[\log \int_{s=0}^{\infty} \frac{D_{t+s} M_{t+s}}{M_t} ds \mid \theta_t\right] & \text{(A.49)} \\ &= \mathbb{E}\left[\log \int_{s=0}^{\infty} D_{t+s} u'\left(C_{t+s}\right) ds \mid \theta_t\right] - \mathbb{E}\left[\log u'\left(C_t\right) \mid \theta_t\right] & \text{(A.50)} \end{aligned}$$

Taking the log of (A.48),

$$\log P_t = \log \mathbb{E}\left[\int_{j=0}^{\infty} u'\left(C_{t+j}\right) D_{t+j} dj \mid Y^t\right] - \log \mathbb{E}\left[u'\left(C_t\right) \mid Y^t\right] \tag{A.51}$$

As in the main text, passing the log through the expectation yields

$$\begin{aligned} p_t &= \mathbb{E}\left[\log \int_{j=0}^{\infty} u'\left(C_{t+j}\right) D_{t+j} dj \mid Y^t\right] - \mathbb{E}\left[\log u'\left(C_t\right) \mid Y^t\right] & \text{(A.52)} \\ &= \mathbb{E}\left[x_t \mid Y^t\right] & \text{(A.53)} \end{aligned}$$

A.4.2 General setup

For a general model, we retain equation (1),

$$P_t = \mathbb{E}\left[\int_0^{\infty} \frac{M_{t+s} D_{t+s} ds}{M_t} \mid \mathcal{F}_t\right] \tag{A.54}$$

where $\mathcal{F}$ here represents the natural filtration induced by Y^t (i.e. $\mathcal{F}_t$ is the sigma-field induced by Y^t). Instead of the θ notation, for consistency here, denote a second filtration, $\mathcal{G}_t$, such that $\mathcal{F}_t$ is coarser than $\mathcal{G}_t$. Then define

$$\begin{aligned} X_t &\equiv \mathbb{E}\left[\int_0^{\infty} \frac{M_{t+s} D_{t+s} ds}{M_t} \mid \mathcal{G}_t\right] & \text{(A.55)} \\ dY_t &= X_t dt + \sigma_{Y,t} dW_t & \text{(A.56)} \end{aligned}$$

and set M_t to be the equilibrium state price process given the filtration $\mathcal{F}_t$. Note that in general models state prices may depend on information, either via endogenous consumption decisions or because information itself affects marginal utility (as in recursive preferences, for example).

We then have, by the law of iterated expectations and the assumption that $\mathcal{F}_t$ is coarser than $\mathcal{G}_t$,

$$P_t = \mathbb{E}\left[X_t \mid \mathcal{F}_t\right] \tag{A.57}$$

as in the main text. The equations, however, define a fixed point – $\mathcal{F}$ depends on Y, which depends on X, which depends on M, which depends on $\mathcal{F}$.

Such a fixed point obviously need not necessarily exist or be unique in any given setting. The analysis in the paper is not meant to fully specify the model but rather to essentially study properties of any model that happens to satisfy (A.57) where $\mathcal{F}$ is the filtration induced by the Y in (A.56).

A.5 Numerical analysis

A.5.1 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.

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.5.2 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.