

Learning and the emergence of nonlinearity in financial markets

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

September 27, 2026

Abstract

Outside the knife-edge case of normally distributed priors, when agents are learning about a hidden state, the higher moments of their beliefs vary endogenously in response to observed signals. This paper formalizes that variation and shows that it is characterized by a simple recursion. Empirically, financial markets are a particularly useful laboratory for studying the model’s predictions because high-frequency data is available along with well-developed methods for estimating higher moments. The predictions of a filtering model for the joint dynamics of the moments of beliefs fit well with a wide range of observed behaviors in financial markets, including, among others, the skewness of returns across frequencies and the volatility, skewness, and persistence of the return volatility process.

1 Introduction

Stock market returns are far from normally distributed. The most salient deviation is that the market sometimes crashes, skewing the distribution of returns left. But returns are much more complicated than simply being independent draws from a negatively skewed distribution. To start, their volatility fluctuates over time, and those fluctuations have been the subject of large literatures in both macroeconomics and finance.¹ Most notably, movements in volatility have a strong negative correlation with market returns (about

*Dew-Becker: Federal Reserve Bank of Chicago; Giglio: Yale University; Molavi: Northwestern University. The views in this paper do not necessarily represent those of the Federal Reserve Bank of Chicago. We appreciate helpful comments and discussions from Marios Angeletos, Rui Da, Jesse Davis, Sergei Glebkin, Jiantao Huang, Paymon Khorrami, Lars Lochstoer, Jean-Paul Renne, Larry Schmidt, and numerous conference and seminar participants.

¹E.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 \(2021\)](#).

–80%) known as the **leverage effect**.² The strength of that relationship changes over time and is related to the conditional skewness of market returns (Neuberger (2012)).³ And volatility itself has nonlinear dynamics, showing positive skewness, time-varying volatility of its own (which is also negatively correlated with returns), and long memory (slowly decaying autocorrelations).⁴ The stock market therefore presents a very rich set of facts for a model to attempt to match.

A large literature studies how investor learning can explain specific individual phenomena in stock returns, such as the leverage effect (Veronesi (1999)) and time-varying skewness (Kozlowski, Veldkamp and Venkateswaran (2020)).⁵ This paper studies learning from a broad and general perspective and argues that it can simultaneously explain a very wide array of behaviors in higher moments of stock returns both qualitatively and quantitatively, wider than perhaps previously recognized. At a more general level, the paper gives a description of the evolution of the moments of agents’ beliefs that applies in broad settings, including in many behavioral models.

Methods

The core of the paper’s theoretical structure is that the net present value (NPV) of a security follows some arbitrary process, and agents continuously receive signals about it. Asset prices in the model are the agents’ best guess of the NPV, which is the solution to a filtering problem. Since the NPV process is essentially unconstrained, the analysis nests a wide range of specifications that have been studied in the literature.

The paper’s key theoretical tool is a novel theorem 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; the sensitivity of the posterior variance is equal to the posterior third moment times signal precision; and so on. The theorem is sufficiently general to also allow feedback from beliefs to fundamentals (e.g. Lorenzoni (2009)).

Results

²See 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.

³For recent work on skewness and the macroeconomy, see Salgado, Guvenen and Bloom (2020), Gormsen and Jensen (2023), Iseringhausen, Petrella and Theodoridis (2023), Dew-Becker (2024), and Menkhoff (2025).

⁴E.g. Mandelbrot (1963), Granger (1980), Mandelbrot, Fisher and Calvet (1997), and Cont (2001).

⁵For past work on belief dynamics and asset prices, among many others, see also Weitzman (2007), David and Veronesi (2013), Collin-Dufresne, Johannes and Lochstoer (2016), Gennaioli, Shleifer and Vishny (2015), Farmer, Nakamura and Steinsson (2024), Wachter and Zhu (2023), and Orlik and Veldkamp (2024).

The paper has three broad classes of results: theoretically deriving a link between unobservable moments of beliefs and observable moments of returns; documenting the model’s quantitative behavior in a simulation; and formally testing its implications in regressions.

The simplest of the theoretical results is the observation that the conditional volatility of returns is proportional to agents’ uncertainty about fundamentals. That allows for a mapping from the model’s prediction for the joint behavior of first and second moments to an empirical prediction for the joint behavior of returns and volatility. The leverage effect – the negative correlation between returns and changes in volatility – appears in the model if and only if agents have negatively skewed beliefs about fundamentals.

There is a similar result for volatility of volatility. Given the required negative skewness of beliefs, a negative correlation between changes in vol-of-vol and returns – which is also observed in the data – appears if and only if agents’ beliefs about fundamentals have positive excess kurtosis.

An immediate implication of the results is to rule out certain processes for fundamentals. For example, a two-state Markov chain for fundamentals, commonly used to capture good and bad times, implies that beliefs become positively skewed in bad times, which is inconsistent with the observation that the leverage effect appears in every month in our sample except two.

The paper’s second result is to show that a quantitative version of the model can both qualitatively and quantitatively match the behavior of aggregate stock returns. Instead of a two-state setup, we assume fundamentals follow a random walk with negatively skewed innovations – think of a model with negative jumps occurring about once every 15 years (calibrated similarly to the estimates in [Barro and Jin \(2011\)](#)).⁶ The fundamentals process in that case has constant volatility, skewness, volatility-of-volatility, etc., and so without learning it would not match any of the moments that we target. With learning, though, it fits well, generating among other things negative skewness in returns that increases with horizon and time-varying return volatility that is positively skewed, negatively correlated with returns, and displays long memory.

Last, the paper tests the model’s predictions more formally via a set of regressions. For the second-moment equation, the model predicts that changes in volatility depend on the interaction of conditional skewness with realized returns and on lagged squared volatility. Both implications hold in the data, and there is a similar equation for changes in the third moment.

⁶For empirical work on skewness in the real economy, see, e.g., [Sichel \(1993\)](#).

The model also implies a restriction across the second- and third-moment equations that their mean-reversion coefficients should be equal, both being determined by the same structural parameter. The regressions show that restriction is consistent with the data. And the value of the structural parameter that they identify is itself consistent with the value that works well in the simulations.

Most of the analysis studies the aggregate US stock market, for which returns are negatively skewed. As an additional test of the model, which need not apply only to stocks, we estimate similar regressions for natural gas futures, whose returns are strongly positively skewed. The second-moment regression results continue to hold, with natural gas also displaying quadratic mean reversion and the sensitivity of its volatility to price changes depending on its conditional third moment.

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 past work on learning has highlighted some of these mechanisms, this paper brings them into a unified framework and gives necessary and sufficient conditions for them to arise. The generality of the analysis also leads to a much wider range of predictions, especially for higher moments, than have been studied in past work.

Additional related work

This paper provides unified results on the relationship between signals and beliefs. That makes it related to work on how uncertainty and responsiveness to news vary endogenously over time, such as [van Nieuwerburgh and Veldkamp \(2010\)](#), [Ludvigson, Ma and Ng \(2021\)](#), [Ilut and Saijo \(2021\)](#), [Altig, Barrero, Bloom, Davis, Meyer and Parker \(2022\)](#), and [Bachmann, Carstensen, Lautenbacher, Menkhoff and Schneider \(2024\)](#). There is a large literature on filtering and expectations formation in economics including, with varying degrees of rationality, [Eusepi and Preston \(2011\)](#), [Nimark \(2014\)](#), and [Bianchi, Ilut and Saijo \(2025\)](#). Our results on long memory are closely related to work on slow convergence to the rational expectations equilibrium.⁷

More directly, the analysis builds on models of asset pricing under learning including [Veronesi \(1999\)](#), [David and Veronesi \(2013\)](#), and [Kozlowski, Veldkamp and Venkateswaran \(2018\)](#). The first two papers study learning about states, while the last is about learning

⁷For example, the methods developed here give a relatively direct and simple derivation of the results in [Christiano, Eichenbaum and Johannsen \(2026\)](#) while simultaneously requiring weaker assumptions about priors.

about time-invariant parameters, but both types of learning are accommodated within this paper’s setup.

Beyond the usual strict Bayesian setup, section 5.1 shows that the paper’s analysis is also 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 \(2024\)](#)) or errors in incorporating new data (e.g. [Bordalo, Gennaioli, Porta and Shleifer \(2019\)](#)).

The quantitative example the paper studies is closely related to work on rare disasters (e.g. [Rietz \(1988\)](#) and [Barro \(2006\)](#)). Even though the probability of a disaster (here just a jump in the fundamental value of stocks) is constant, at any given time agents are unsure whether a disaster has occurred, so their subjective distribution over future returns varies over time *as though* the probability of a disaster is time-varying, as in [Gabaix \(2012\)](#) and [Wachter \(2013\)](#) (see also [Ghaderi, Kilic and Seo \(2022\)](#)).⁸

Finally, there are obviously other explanations in the literature for some of the phenomena studied in the paper. A common explanation for the leverage effect is that it comes from shocks to uncertainty or volatility raising expected returns and hence reducing prices (e.g. [Campbell and Hentschel \(1992\)](#) and [Bekaert and Wu \(2000\)](#)). This paper’s framework is general enough to allow that mechanism, but it primarily focuses on effects coming purely through learning.

Perhaps the strongest distinction between this paper’s analysis and that in past work is the tight structure on higher moments of beliefs and returns. Models of volatility feedback, for example, are silent on the higher-order properties of volatility, such as vol-of-vol and skewness of volatility. The model here delivers strong implications for how those moments should behave and finds support for those predictions in the data. Similarly, unlike this paper, other work does not get long memory in volatility endogenously.

2 Motivating facts

We begin with a set of motivating facts.⁹ All moments for realized returns are for the S&P 500 total return (from CRSP). The calculation of conditional moments is described in appendix E; it uses information from option markets as well as other predictors.

Table 1 reports moments of daily market returns, their conditional volatility, and the daily change in that conditional volatility. Returns are slightly negatively skewed, while

⁸See also [Wachter and Zhu \(2023\)](#) for a model of learning with rare disasters and [Maenhout, Xing and Balter \(2025\)](#) for related work with ambiguity aversion.

⁹[Cont \(2001\)](#) notes that similar behavior is observed across many different financial markets.

Table 1: Stock market return and volatility moments

Moment	Stock market	Volatility	
	Daily return	Level	Daily change
Std. dev.	1.14	6.61	1.42
Skewness	−0.06	2.19	1.35
Kurtosis	19.56	11.60	29.96
Corr. w/ R_t			−0.79

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 and is reported in annualized standard deviation units.

volatility is highly positively skewed in both levels and changes. All three series have severe excess kurtosis.

The top-left panel of figure 1 plots the skewness of realized log returns over holding periods of 1 to 252 trading days. Skewness becomes significantly more negative as the horizon increases. It is -0.39 at the daily level and -1.24 at a horizon of a year. The top-right panel plots the time series of conditional skewness for one-month market returns (i.e. the time series of the 21st point in the top-left panel). It has been negative through the entire sample.

The middle-left panel of figure 1 is a scatter plot of daily market returns against the daily change in their volatility, showing the strong negative correlation referred to as the leverage effect. The correlation coefficient in the sample is -0.79 . The middle-right panel plots estimates of the regression coefficient in every month between 1990 and 2026, 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, similar to market skewness in the top-right panel.

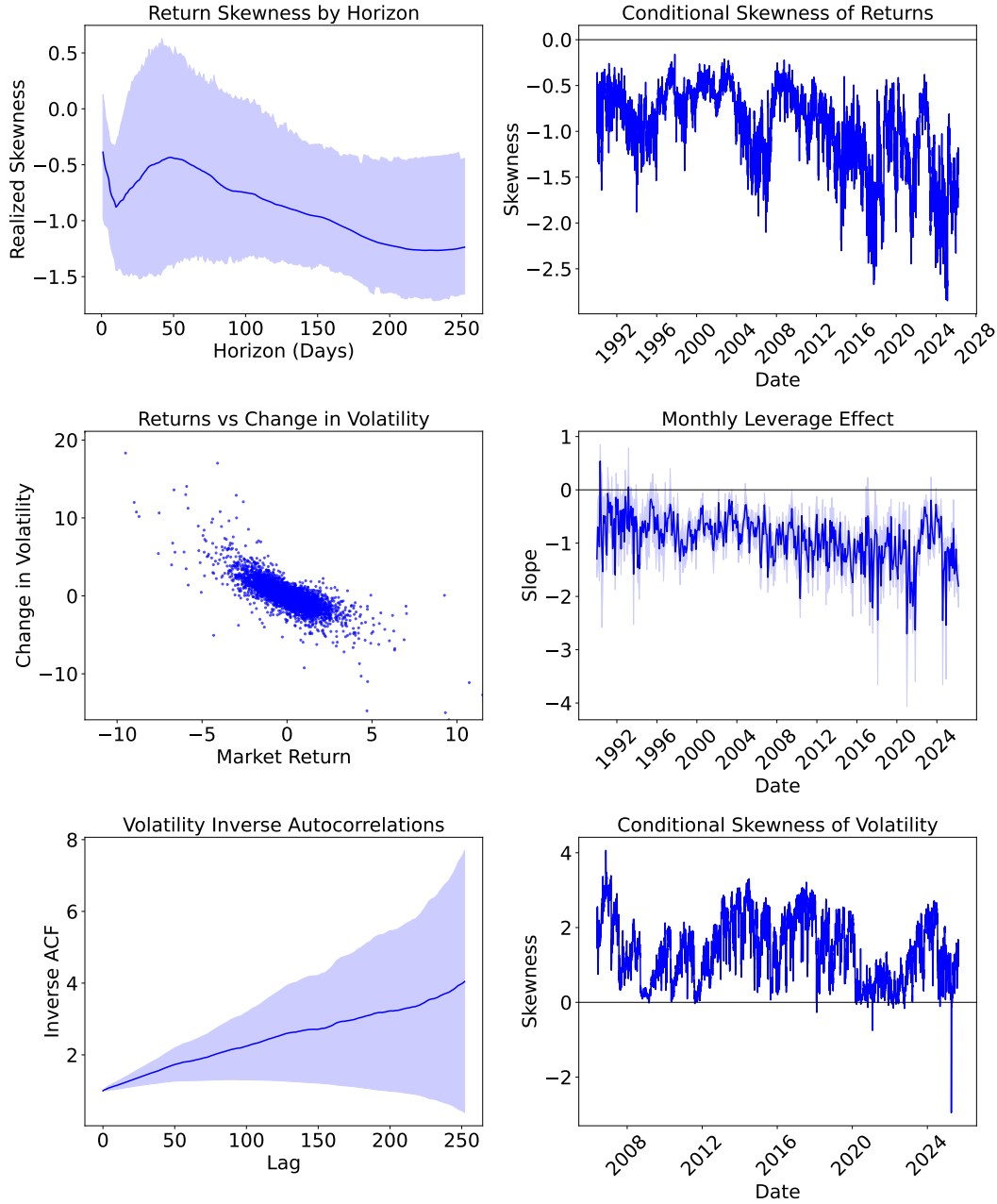
Finally, the bottom panels of figure 1 plot moments for conditional volatility. The bottom-left panel displays its inverse autocorrelations, $1/\text{corr}(\text{vol}_t, \text{vol}_{t-j})$.

Inverting the autocorrelations helps to visualize the deviation from the exponential decay that would be expected if volatility followed an ARMA process. The fact that $1/\text{corr}(\text{vol}_t, \text{vol}_{t-j})$ grows approximately linearly is consistent with polynomial decay in the autocorrelations, as in fractionally integrated models like [Ding, Granger and Engle \(1993\)](#) and [Bollerslev and Mikkelsen \(1996\)](#).¹⁰

The bottom-right panel plots the time series of conditional skewness of conditional

¹⁰Though note that fractional integration is an asymptotic concept and there are infinitely many ARMA models that can match any finite set of autocorrelations.

Figure 1: Motivating evidence



Note: All panels use daily data from 1990, except the top-left panel (daily returns from 1926) and the bottom-right panel (VIX-option data available from May 2006 through August 2025). Shaded areas are 95% confidence intervals. Construction of the conditional moments is described in appendix E.

volatility at the one-month horizon, $skew_t(vol_{t+21})$. That skewness has been positive on almost every day in the sample.

The paper's aim is to qualitatively and quantitatively understand all of these moments, both conditionally and unconditionally. Being able to do so in a parsimonious framework with limited assumptions is what distinguishes this paper's model from many alternatives.

3 Model setup and solution

3.1 Model setup

Stocks pay some cash-flow D_t and there is a stochastic discount factor M_t that prices D_t . We denote the information that agents have on date t , which will be described in more detail below, by Y^t . The price of stocks is, definitionally,

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

where $\mathbb{E}$ is the expectation operator. Importantly, D and M may depend on Y . Since M is left unconstrained, the model allows for unrestricted risk premia, which may vary over time.

3.1.1 Information

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

$$X_t \equiv \mathbb{E} \left[\int_{s=0}^{\infty} \frac{D_{t+s} M_{t+s}}{M_t} ds \mid \theta_t \right]. \quad (2)$$

That is, X_t is just the NPV for a hypothetical investor with complete knowledge. An extreme case is perfect foresight, in which θ_t contains full information of all values of M and D in the future, but θ_t can also be coarser. While X_t itself is a scalar, it is a function of θ_t , meaning its dynamics can be very rich and can accommodate, for example, many forms of nonstationarity.

The law of iterated expectations immediately implies

$$P_t = \mathbb{E} [X_t \mid Y^t]. \quad (3)$$

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

Assumption 1 *In the theoretical analysis, the log price is*

$$p_t \equiv \mathbb{E} [x_t \mid Y^t], \quad (4)$$

where x_t is the log NPV process,

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

That act of passing the log through the expectation is the single approximation step in the theoretical analysis. Similar transformations are common – analyses of macro-finance models often rely on the Campbell–Shiller approximation, for example. Appendix C 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 at all.

The key assumption giving the model economic content is the specification of agents’ information.

Assumption 2 *Agents’ information is summarized by a sufficient statistic Y satisfying*

$$dY_t = x_t dt + \sigma_Y dW_t, \quad (6)$$

where W is a standard Brownian motion.

Assumption 2 can accommodate different underlying information structures. Our preferred interpretation is that Y_t summarizes the content of the broad range of sources investors have access to. That will hold (1) if there are a finite number of those sources, all Gaussian (since then they always have a sufficient statistic of the form (6)) or (2) if there are many individual sources of information, not necessarily Gaussian, but which are all asymptotically only weakly informative. When such signals are combined, they take the form (6) via a martingale central limit theorem. In a richer model, Y_t might be an endogenous outcome of an information processing or rational inattention type of problem.

Given that x_t represents how agents would value stocks if they had complete information, it makes a certain amount of sense to assume that it is what agents learn about, but that is

not the only possible setup. Agents could instead 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 . 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_t , but pays for that generality with this restriction on the information structure. Given that we do not know exactly what information investors observe (e.g. see Roll (1988)), the reasonableness of assumption 2 should be judged based on its observable predictions for the behavior of endogenous outcomes like prices.

Beyond the substantive assumption on the information structure, we assume that x_t satisfies the technical regularity conditions stated in the appendix. These assumptions are weak and are satisfied by essentially any stochastic process used in economics.

3.1.2 Feedback from information to fundamentals

A significant potential complication here is the possibility of feedback from beliefs to fundamentals. The most common source of feedback is that state prices can depend on beliefs, as under recursive preferences and ambiguity aversion, or where state prices depend on the level of market wealth, as in the CAPM. But it is also possible for cash flows to depend on beliefs – there is a large literature in macroeconomics allowing that sort of feedback.

Formally, the feedback in the model is measured by the covariance of x with errors in the signal, normalized by the usual Bayesian learning gain:¹¹

$$a_t \equiv \frac{\sigma_Y}{\text{var}_t(x_t)} \frac{\text{cov}_{t-}(x_t, dW_t)}{dt}. \quad (7)$$

If fundamentals are strictly exogenous, as in an endowment economy with time-separable preferences, then a_t is identically zero. But when the noise investors observe can have real effects, a_t becomes nonzero.

Proposition 1 $a_t = 0$ for all t if either of the following is true:

- a. The discount rate (i.e. sum of the risk-free rate and risk premium) is constant and D_t is exogenous to W_t .
- b. The (D_t, M_t) pair is exogenous to W_t , as in the endowment economy of appendix D.

¹¹Formally, $\text{cov}_{t-}(x_t, dW_t)/dt = \text{cov}_{t-}(dx_t, dW_t)/dt = d\langle x, W \rangle_t/dt$, where $\langle x, W \rangle$ is the predictable quadratic covariation of x and W .

For simplicity, we derive our results under the assumption that a_t is Y^t -measurable.

3.2 Solution

Since the log price is the expectation of x_t , the return depends on the change in the conditional expectation of x_t , which takes a form very similar to what is obtained under the Kalman filter.

Proposition 2 *In the structure set up so far,*

$$dp_t = \kappa_{2,t} (1 + a_t) \sigma_Y^{-2} (dY_t - \mathbb{E}_t [x_t] dt) + \mathbb{E}_t [dp_t], \quad (8)$$

where $\kappa_{2,t} \equiv \text{var}_t (x_t)$.

The sensitivity of prices to surprises in the signal, $dY_t - \mathbb{E}_t [x_t] dt$, depends on current uncertainty about fundamentals ($\kappa_{2,t}$), signal precision (σ_Y^{-2}), and the relative strength of the feedback effect (a_t). The intuition for the gain is simple: it is the coefficient from a hypothetical local regression of x_t on dY_t . The drift, $\mathbb{E}_t [dp_t]$, captures the risk-free rate and the expected risk premium.

Proposition 2 *[continued...]* Denote the posterior third central moment by $\kappa_{3,t}$. Then

$$\begin{aligned} d\kappa_{2,t} = & \kappa_{3,t} \sigma_Y^{-2} (dY_t - \mathbb{E}_t [x_t] dt) - \kappa_{2,t}^2 (1 + a_t)^2 \sigma_Y^{-2} dt \\ & + \text{var}_t (dx_t) + 2 \text{cov}_t (x_{t-}, dx_t). \end{aligned} \quad (9)$$

The dynamics of the variance $\kappa_{2,t}$, which determines return volatility, are similar to those for prices. The gain of $\kappa_{2,t}$ is now the *third* moment times the precision of the signal (with no adjustment term). That is again because $\kappa_{3,t}/\sigma_Y^2$ is a local regression coefficient. $\text{var}_t (dx_t)$ represents the contribution of uncertainty from shocks to fundamentals, and $2 \text{cov}_t (x_{t-}, dx_t)$ is the contribution due to x effectively spreading out over time.

Proposition 2 *[continued...]*

$$\begin{aligned} d\kappa_{3,t} = & \kappa_{4,t} \sigma_Y^{-2} (dY_t - \mathbb{E}_t [x_t] dt) - 3\kappa_{2,t} (1 + a_t) \sigma_Y^{-2} \kappa_{3,t} dt \\ & + 3 (\mathbb{E}_t [x_t]^2 - \kappa_{2,t}) \mathbb{E}_t [dx_t] - 3\mathbb{E}_t [x_t] \mathbb{E}_t [d(x_t^2)] + \mathbb{E}_t [d(x_t^3)], \end{aligned} \quad (10)$$

where $\kappa_{4,t}$ is the fourth **cumulant** (equal to zero under the normal distribution).

For the third moment, the sensitivity to signals now depends on the fourth cumulant. We will show below that the determinants of the third moment matter because they determine the sign of skewness in volatility and also affect fragility in asset prices. $\kappa_{4,t} > 0$ is equivalent to excess kurtosis being positive and means that the distribution has more mass in the tails than the normal distribution. In that case, good news increases skewness because it shrinks the left tail of beliefs faster than the right.

3.2.1 General result

While proposition 2 is enough for the present paper, it suggests 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 cumulant is the mean, and the second and third cumulants equal the second and third central moments. Denote the n -th cumulant of the time- t conditional distribution of x_t by $\kappa_{n,t}$.

Theorem 1 *For all n for which the $(n+1)$ th posterior cumulant exists and for which the expectations below are well defined,¹²*

$$\begin{aligned} d\kappa_{n,t} = & (\kappa_{n+1,t} + \mathbf{1}_{n=1}\sigma_Y\alpha_t)\sigma_Y^{-2}(dY_t - \mathbb{E}_t[x_t]dt) \\ & - \frac{1}{2\sigma_Y^2}\mathbf{1}_{n\geq 2}\left[\sum_{j=2}^n\binom{n}{j-1}\kappa_{j,t}\kappa_{n-j+2,t} + 2n\sigma_Y\alpha_t\kappa_{n,t} + 2\sigma_Y^2\alpha_t^2\mathbf{1}_{n=2}\right]dt \\ & + \sum_{j=1}^n\binom{n}{j}B_{n-j}(-\kappa_{1,t}, \dots, -\kappa_{n-j,t})\mathbb{E}_t[d(x_t^j)], \end{aligned} \quad (11)$$

where B_j denotes the j th complete exponential Bell polynomial, $\mathbf{1}_{\cdot}$ is the indicator function, and $\alpha_t \equiv \text{cov}_{t-}(x_t, dW_t)/dt$ so that $a_t = \sigma_Y\alpha_t/\kappa_{2,t}$.

The recursion for the gain carries through to all the cumulants: the gain of the n th cumulant is the $(n+1)$ th cumulant times the precision of the signal. Theorem 1 follows from an application of results in Liptser and Shiryaev (2013) and Bain and Crisan (2009).¹³ While the equation for $\kappa_{1,t}$ is well known (equation (8)), the full recursion is not

¹²Since the cumulants are derivatives of a function, if $\kappa_{n+1,t}$ exists then all lower-order cumulants also exist.

¹³Theorem 1 is closely related to results in Dytso, Poor and Shitz (2022), with three key differences. First, x_t here is dynamic instead of constant. Second, here there is feedback from the signals to the latent state via α_t . Third, theorem 1 characterizes the evolution of the conditional cumulants from knowledge only of the priors. In unpublished work, Osband (2011) and Osband (2025) get a special case of theorem 1 when x is a pure jump process and $\alpha_t = 0$ for all t (closely related to Veronesi (1999)).

a trivial implication of the basic filtering equations, especially with feedback from signals to fundamentals.

4 Predictions

This section examines the predictions of proposition 2 for the behavior of returns. We allow the feedback strength, a_t , to vary over time but assume that it changes smoothly. We also assume that the conditional moments of x_t are well behaved.¹⁴

4.1 Volatility and uncertainty

From equation (8) we immediately have for volatility¹⁵

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

$$= \kappa_{2,t} (1 + a_t) \sigma_Y^{-1}. \quad (13)$$

Combining equations (8) and (9) from proposition 2 then yields the following.

Corollary 1 *vol_t follows a diffusion satisfying*

$$\begin{aligned} d(vol_t) = & \underbrace{\sigma_Y^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}} (dp_t - \mathbb{E}_t [dp_t])}_{\textcircled{a}} - \underbrace{\sigma_Y^{-1} (1 + a_t) vol_t^2 dt}_{\textcircled{b}} \\ & + (1 + a_t) \sigma_Y^{-1} (\text{var}_t(dx_t) + 2 \text{cov}_t(x_{t-}, dx_t)) + \sigma_Y^{-1} \kappa_{2,t} da_t. \end{aligned} \quad (14)$$

The main predictions arise out of the terms $\textcircled{a}$ and $\textcircled{b}$. $\textcircled{a}$ determines the joint behavior of returns and volatility, while $\textcircled{b}$ generates nonlinearity in the dynamics of volatility.

A first point to note is that volatility is generically stochastic. It is only when the model is linear Gaussian and feedback strength is constant that the volatility of prices deterministically converges to a constant. If any of the higher-order cumulants is nonzero,

¹⁴Specifically, we assume that the technical conditions of theorem 1 hold through order four, that $1 + a_t > 0$, that a_t is continuously differentiable with a locally bounded rate of change, and that the conditional drift rates $\mathbb{E}_t[d(x_t^j)]/dt$, $j = 1, \dots, 4$, are locally bounded and continuous.

¹⁵For 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 is 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.

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.

4.2 The leverage effect

Proposition 3 *The instantaneous coefficient in a regression of changes in the conditional volatility of returns on price changes is*

$$\frac{\text{cov}_t(dp_t, d\text{vol}_t)}{\text{var}_t(dp_t)} = \sigma_Y^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}}. \quad (15)$$

The term @ in (14) shows that the presence of a **leverage effect** – the negative correlation between changes in volatility and prices documented in table 1 and 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 $\kappa_{3,t} < 0$: there is a leverage effect if and only if agents’ posterior distribution for fundamentals is negatively skewed. The fact that we observe a leverage effect in the aggregate US stock market in nearly all months in the data, including during severe downturns, then implies that the conditional skewness is negative in essentially all states of the world observed in our sample. The relationship has additionally strengthened over time, which would be consistent with a decline in $\kappa_{3,t}$.

The intuition for proposition 3 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.

4.3 Skewness in returns

Since the price process is a diffusion, its instantaneous skewness is not formally well defined. 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:

$$\text{skew}_t(p_{t+\Delta t}) \equiv \frac{\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t p_{t+\Delta t})^3]}{\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t p_{t+\Delta t})^2]^{3/2}}. \quad (16)$$

Proposition 4 *The local skewness of returns is*

$$\lim_{\Delta t \downarrow 0} skew_t(p_{t+\Delta t}) (\Delta t)^{-1/2} = 3 \frac{\kappa_{3,t}}{\kappa_{2,t}} \sigma_Y^{-1}. \quad (17)$$

The conditional “instantaneous” skewness of returns again depends on the second and third moments of the posterior. As $\Delta t \rightarrow 0$, skewness goes to zero – that is the usual result that returns are locally normal. For small but nonzero values of Δt , equation (17) provides a link between the conditional skewness of returns – which is potentially measurable – and the conditional skewness of beliefs, $skew_t(x_t)$, which determines the leverage effect:

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

$$\lim_{\Delta t \downarrow 0} skew_t(p_{t+\Delta t}) \frac{1}{3} (\Delta t)^{-1/2} = \frac{\text{cov}_t(dp_t, dvol_t)}{\text{var}_t(dp_t)}. \quad (18)$$

These results show that the model is able to qualitatively match the return skewness documented in table 1 and figure 1 again when $\kappa_{3,t}$ is negative. Skewness arises here again due to the term ② in equation (14): 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.

Additionally, corollary 2 implies that the conditional skewness and leverage effect plotted separately in figure 1 should actually be proportional to each other. Section 7 tests that proposition.

The skewness of returns is closely linked to the concept of fragility in prices (Taleb and Douady (2013)). Formally, the response of prices to signals in the model is concave when $\kappa_{3,t} < 0$. That concavity then generates negative skewness in returns and a leverage effect. Equation (10) therefore describes the dynamics of fragility in the model. When beliefs are heavy-tailed in the sense of $\kappa_{4,t} > 0$, fragility increases following negative signals about fundamentals.

4.4 Conditional moments of volatility

Starting from corollary 1, the same analysis that leads to equations (13) and (17) yields the volatility and skewness of volatility.

Proposition 5 *The local volatility and skewness of vol_t are*

$$\left(\lim_{\Delta t \downarrow 0} \mathbb{E}_t [(vol_{t+\Delta t} - \mathbb{E}_t vol_{t+\Delta t})^2] / \Delta t \right)^{1/2} = |\kappa_{3,t}| (1 + a_t) \sigma_Y^{-2}, \quad (19)$$

$$\lim_{\Delta t \downarrow 0} skew_t(vol_{t+\Delta t}) (\Delta t)^{-1/2} = 3 \frac{\kappa_{4,t}}{|\kappa_{3,t}|} \sigma_Y^{-1}. \quad (20)$$

Equation (19) says that the volatility of volatility is driven by the third moment of beliefs, which follows simply due to the basic moment recursion in the Bayesian update. Skewness of returns and volatility of volatility are therefore closely related, both being driven by the same feature of beliefs. The feedback term a_t appears since it determines the scale of the volatility process.

The skewness coefficient for volatility takes the same form that we see for return skewness, but with the cumulants shifted back by one order. Similar to the result in the previous section that the negative leverage effect coefficient is equivalent to a negative third moment of beliefs, here positive skewness in volatility – which is documented empirically in table 1 – is equivalent to a positive fourth cumulant, meaning that beliefs have greater kurtosis than the normal distribution. The beta distribution is an example of a distribution that can have a negative fourth cumulant, yielding negative skewness in volatility even with negative skewness in fundamentals.

Proposition 5 shows a way that this model is different from a simple model with time-varying volatility as an exogenous driving process (e.g. [Bansal and Yaron \(2004\)](#)): volatility is not a free process here. It has rich dynamics that are constrained by the endogenous dynamics of beliefs.

4.5 Slow decay in volatility

The term ⑥ in (14) shows how volatility decays. When volatility is high, $\sigma_Y^{-1} (1 + a_t) vol_t^2$ also grows, pulling volatility back down towards its steady state. The rate at which volatility decays following an increase depends on $\sigma_Y^{-1} (1 + a_t)$. The rate of decay is increasing in the precision of signals, while feedback from beliefs to fundamentals via the term a_t accelerates the recovery of volatility (holding the conditional variance of changes in fundamentals fixed) because it increases the covariance of signals with fundamentals.

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. 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 $1/(1 + b\Delta t)$

for a coefficient b .¹⁶ The polynomial decay is consistent both with the autocorrelations plotted in figure 1 and also with past empirical work on securities markets.¹⁷

Intuitively, volatility decays nonlinearly because the degree to which expectations 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. The same mechanism is what drives the $1/N$ rate of convergence for the variance of statistical estimators.

4.6 Examples

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

4.6.1 Linear Gaussian process

If fundamentals, x_t , follow a linear Gaussian process and the strength of feedback, a_t , is zero, 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 cumulants are always equal to zero. There is then no leverage effect, volatility of volatility, or skewness in prices or volatility.

4.6.2 Markov switching process

Veronesi (1999) studies a two-state switching model in which the latent state x switches between a low and a high value and agents have a Gaussian signal as required in proposition 2. In this case, the low and high values of x_t can be normalized to 0 and 1 without loss of generality.

Agents' beliefs at any given time are summarized by a single parameter, π_t , which is their posterior probability that $x_t = 1$. The cumulants of x_t , which drive price dynamics, are simple functions of π_t :

$$\kappa_{2,t} = \pi_t (1 - \pi_t), \quad (21)$$

$$\kappa_{3,t} = (1 - 2\pi_t) \times \kappa_{2,t}, \quad (22)$$

$$\kappa_{4,t} = \kappa_{2,t} (1 - 6\kappa_{2,t}). \quad (23)$$

¹⁶Specifically, if there is a jump at some date t_0 , then if there are no further shocks to volatility and a_t is constant, then to leading order $vol_{t_0+\Delta t} \approx vol_{t_0} / (1 + \sigma_Y^{-1} (1 + a_{t_0}) vol_{t_0} \Delta t)$.

¹⁷See 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 always consistent both with strict long memory and also processes that simply approximate it, since formally long memory is defined asymptotically.

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 $\text{sign}(\kappa_{3,t}) = \text{sign}(\frac{1}{2} - \pi_t)$. Economically, when π_t is near 1 so that agents are confident they are in the good state, volatility is low, but the skewness coefficient 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 $1/2$, volatility falls and the leverage effect *reverses*: agents no longer worry as much about the economy getting worse, so there is relatively more upside and $\kappa_{3,t} > 0$. That fact illustrates 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.

Similarly, the fourth cumulant can be shown to change sign twice – it is positive for values of π close to 0 and 1 and negative for intermediate values. Empirically, figure 1 shows that the conditional skewness of volatility, whose sign is the same as that of $\kappa_{4,t}$, is uniformly positive in the data, which is not possible in a two-state model.

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 in the simulations below, the analysis does not require it. The recursive structure of proposition 2 holds as long as agents update using the basic *form* of Bayes’ rule. In particular, none of the following is essential in the derivations:

1. That agents use the correct precision (σ_Y^{-2}) in calculating their update (e.g. the diagnostic expectations of [Bordalo, Gennaioli, Porta and Shleifer \(2019\)](#)). If the perceived precision is incorrect, the perceived precision replaces σ_Y^{-2} in the gain terms of proposition 2 and the drift terms acquire an additional correction that is proportional to the error in the perceived signal variance.
2. That agents’ assumed dynamics for x are in any sense “correct.”
3. That the agents incorporate all information they receive. They could irrationally or inefficiently ignore some signals.

4. 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.

What is important here is not actually that agents are true Bayesians but that their updating rule has the algebraic structure of Bayes' rule. Deviations from rational expectations will mean that what agents see as a martingale will not appear to be one to an econometrician, so the drift terms in the results above inherit agents' subjective model. But the nonlinearities the paper focuses on do not hinge on anything like complete rationality. [Molavi \(2026\)](#) discusses these issues in greater detail.

5.2 Discrete information revelation events

[Dytso, Poor and Shitz \(2022\)](#) precedes this paper and proves the following discrete version of theorem 1, under the additional restriction of a constant latent state (and hence no feedback from signals to fundamentals). Instead of assuming a diffusive information flow, this result is for a signal with strictly positive information content, meaning that the moments also update by discrete amounts.

Proposition 6 [[Dytso, Poor and Shitz \(2022\)](#), equation (52)] *For a random variable x_t and a signal $y_t \sim N(x_t, \sigma^2)$ (with $y_t - x_t$ independent of x_t)*

$$\frac{d}{dy} \kappa_j(x_t | y_t = \bar{y}) = \kappa_{j+1}(x_t | y_t = \bar{y}) / \sigma^2, \quad (24)$$

where $\kappa_j(x_t | y_t = \bar{y})$ is the j th posterior cumulant of x_t conditional on observing $y_t = \bar{y}$. Furthermore, for y_t in a neighborhood of any $\bar{y} \in \mathbb{R}$,

$$\mathbb{E}(x_t | y_t) = \sum_{j=0}^{\infty} \frac{\kappa_{j+1}(x_t | y_t = \bar{y})}{j!} \left(\frac{y_t - \bar{y}}{\sigma^2} \right)^j. \quad (25)$$

Proposition 6 shows that the type of recursion in theorem 1 also holds for discrete revelation events – diffusive information is not necessary for the central results. At the same time, it shows that normality is important – continuity is not essential, but proposition 6 still requires a normally distributed signal.¹⁸ Proposition 6 also shows why continuous time is useful here: in (24) the *posterior* cumulants, which are precisely what

¹⁸[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.

one wants to solve for, determine sensitivity. In continuous time the cumulants follow continuous processes, so the prior and posterior values are effectively identical, which simplifies the analysis.

Proposition 6 is closely related to and can be derived from theorem 1. Theorem 1, on the other hand, does not follow directly from proposition 6. Accounting for dynamics, which is ultimately central to the analysis, as well as feedback from signals to fundamentals, which are not addressed in Dytso, Poor and Shitz (2022), makes the derivations significantly more complicated.

6 Illustrative calibration

This section develops a simple quantitative example to help understand the extent to which the qualitative theoretical predictions map into quantitatively realistic behavior.

6.1 Model setup

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

$$dx_t = (\phi\lambda + g) dt - J_t dN_t + \sigma_x dB_t, \quad (26)$$

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 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. Feedback from signals to fundamentals is turned off here, so $a_t = 0$.

In the absence 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 and skewness would be constant, there would be no leverage effect or long memory, and volatility of volatility and skewness of volatility would be equal to zero. The version of the model without learning would match none of the behavior observed in figure 1 and little of what is in table 1.

We calibrate the jump distribution based on Barro and Jin (2011) (though with smaller disasters assumed to have somewhat heavier tails) and assume that equities have a leverage of 3 relative to shocks to consumption.¹⁹ σ_x is then chosen to approximately match

¹⁹Leverage of 3 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).

the standard deviation of US stock returns ($\sigma_x = 0.0038$ at the daily frequency), and $\sigma_Y = 4.5$ (based on the time index corresponding to days) was chosen to match the dynamics of volatility and skewness along with regression results reported in section 7. 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 0.28 in logs. Since σ_Y controls the rate of learning, it determines both the autocorrelations of volatility and also the time from peak to trough following downward jumps in fundamentals. We compare both to the data below to help validate the choice of σ_Y . Finally, we set $g = 0.07$ to match the historical equity premium.

It is straightforward to simulate the model by discretizing the state space and then updating beliefs exactly using Bayes' theorem. 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, $J_t dN_t$.

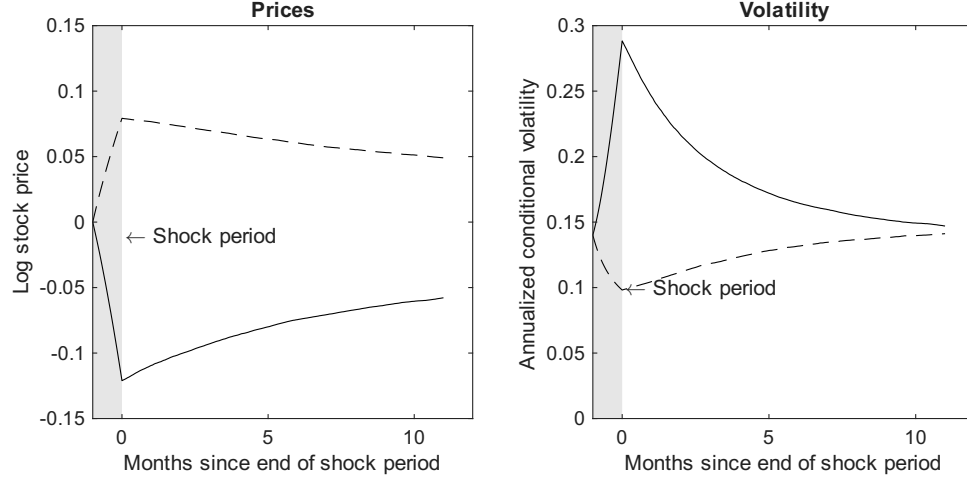
The noise shock is defined as a month with a one-in-ten-year (± 2.4 standard deviation) realization in the total error in the signal over the month. Figure 2 plots the response of prices and volatility to the shock. The first month in the figure is the period in which the shock occurs. In the case of a negative shock, 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 (a closely related mechanism underlies the results of [Christiano, Eichenbaum and Johannsen \(2026\)](#)).

In the case of a positive noise shock, the effects are smaller because of the asymmetry in the fundamentals process. Agents understand that it is possible that a large negative

The exact calibration is the double power law with, in their notation, $\alpha = 4.16$, $\beta = 5$, $\delta = 1.38$, with minimum and maximum consumption jump sizes of 7.4 and 65.6 percent, respectively, with the maximum cutoff based on the largest observed decline in their data. The annual jump probability is 6.42 percent.

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 shock is on average 54% 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



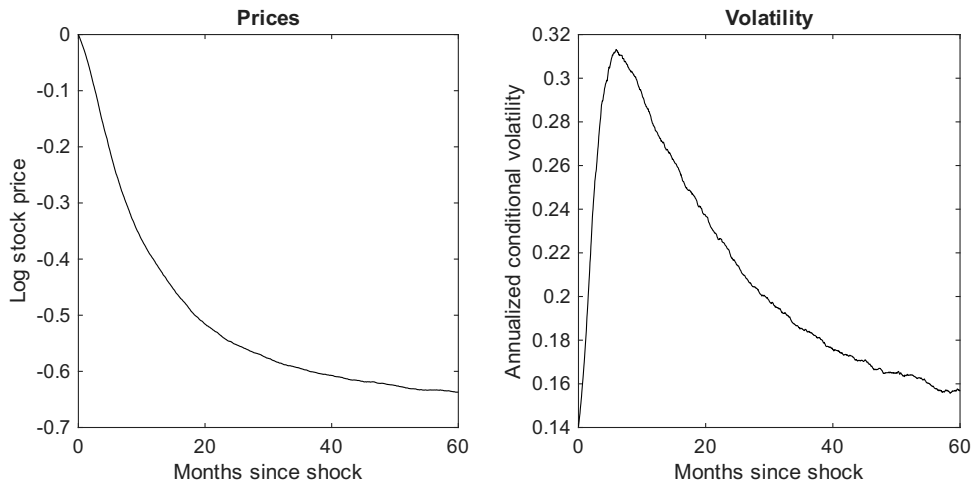
Note: The left-hand panel plots the IRF for prices – the conditional expectation of x – and the right-hand panel plots the conditional standard deviation of prices. The shock is a one-time 2.4 standard deviation negative error in the signal (i.e. a negative realization of $\sigma_Y dW$).

The second pair of IRFs, 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 0.14 percent per day, accelerating to 0.21 percent per day at their peak. It takes four to 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. This behavior is similar to that studied in Ghaderi, Kilic and Seo (2022), with the difference here being that it happens endogenously, instead of there being an assumption that disaster shocks are clustered.

To get a sense of whether the length of crashes is empirically reasonable, figure E.2 in the appendix plots the time from peak to trough for all drawdowns in the US stock market greater than 10 percent compared to the average time for drawdowns of the same magnitude in the simulations. If anything, the model actually appears to *understate* the time from peak to trough relative to the data, implying that learning in the data is even slower than what is calibrated here.

The right-hand panel of figure 3 shows that the peak in volatility does not come on average until 9 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 84 percent, nearly twice as high as the average five-year peak of only 44 percent for periods with no jump.

Figure 3: Response to a negative realization of fundamentals



Note: These plots are the same as in figure 2, 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.

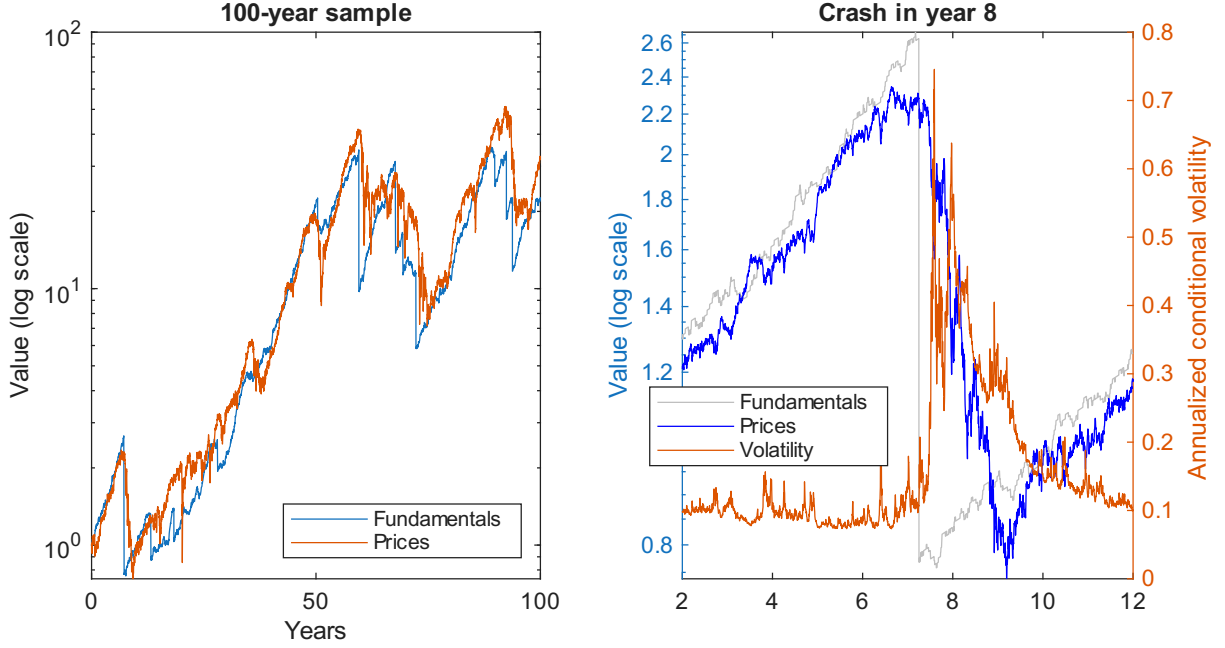
6.3 Simulation results

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. In some cases, fundamentals jump down and it takes time for prices to catch up, and there are also clear cases of prices jumping down “erroneously” (based on hindsight or on knowing the true state) and then recovering, 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, while volatility rises as prices fall.

Table 2 reports moments for returns and their conditional volatility. As discussed in

Figure 4: Simulated time series of fundamentals and prices



Note: “Fundamentals” is the simulated x process, and “prices” is the log of the simulated price process.

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 worse than the most negative value observed in the US over the past century. Calculating moments on 100-year samples also accounts for any small-sample bias in the statistics.

Table 2 shows that the model broadly matches the data. The moments for daily returns all lie within the 95% confidence bands of the empirical estimates, as do the standard deviations of the level and changes of volatility. The model’s correlation between volatility changes and returns is somewhat stronger than in the data (-0.86 versus -0.79), though the empirical value lies well within the range of correlations across individual simulated samples. The model overstates the skewness and kurtosis of the level of volatility relative to the data, while for first differences it understates skewness and overstates kurtosis.

Figure 5 compares the model’s behavior to what was shown in figure 1. Log return skewness and volatility autocorrelations are both qualitatively similar to the behavior of the data across horizons.²⁰ The bottom-left panel shows that the model generates a

²⁰Note that the value of skewness differs here from table 1 because the figure uses log returns while the table uses level returns. The reason is that at long horizons, we want skewness to converge to zero, consistent with the CLT. That only works for log returns. Due to compounding, level returns are

Table 2: Model vs data moments

Moment	Stock returns		Volatility level		Volatility change	
	Data	Model	Data	Model	Data	Model
Std. dev.	1.14	1.00	6.61	6.53	1.42	1.21
	[1.00, 1.28]		[5.02, 8.26]		[1.19, 1.66]	
Skewness	−0.06	0.01	2.19	3.63	1.35	0.06
	[−0.71, 0.42]		[1.00, 2.80]		[0.58, 2.24]	
Kurtosis	19.56	20.23	11.60	26.34	29.96	93.25
	[13.76, 27.22]		[3.87, 15.95]		[17.98, 41.76]	
Corr. w/ R_t					−0.79	−0.86
					[−0.82, −0.76]	

Note: This replicates table 1 for the empirical moments and compares them to the model simulation. Square brackets below the data moments are 95% confidence intervals from a block bootstrap of the data (252-day blocks, 1,000 replications).

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 densities for volatility in the model and data demonstrating that the distributions are reasonably similar, though, as is also reflected in table 2, the skewness in the simulation is much higher than the data.

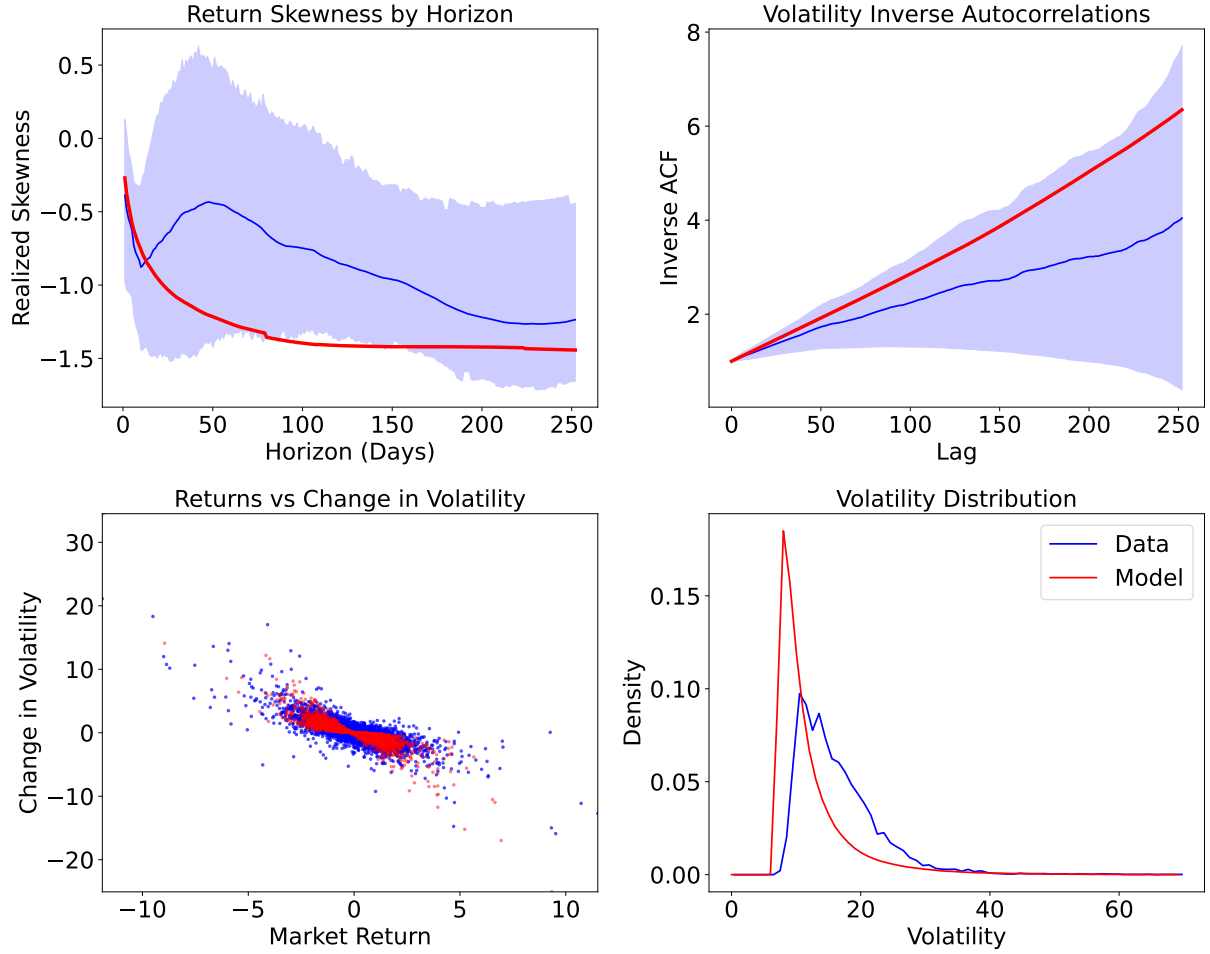
The results in this section show that the model is able to match key features of the data not just qualitatively but also quantitatively, and with only a relatively small number of free parameters. The overall volatility of returns and the fact that they are negatively skewed come from the calibration of fundamentals. But the volatility dynamics and the evolution of skewness across horizons are both absent from fundamentals and arise from the learning mechanism.

7 Estimated belief moment dynamics

This section estimates regressions motivated by the model-implied dynamics for the second and third moments of beliefs from section 4. As a simple benchmark, we assume that x has stationary and independent increments and set $a_t = 0$.

asymptotically log-normal and therefore positively skewed. The theory is also entirely about log returns.

Figure 5: Simulation results



Note: The top two panels and the bottom-left panel repeat the return-skewness, volatility inverse-autocorrelation and return-volatility scatter panels of figure 1, overlaid with the corresponding outputs from simulations 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.

7.1 Regression setup

7.1.1 The κ_2 equation

Combining equations (14) and (17) and assuming $\mathbb{E}_t[dp_t]$ is zero for simplicity,²¹ we have

$$d[vol_t] = cons + \frac{1}{3}skew_t dp_t - \sigma_Y^{-1} vol_t^2 dt, \quad (27)$$

²¹The results of this section are essentially identical if we replace dp_t with $dp_t - \mathbb{E}_t[dp_t]$ and assume that $\mathbb{E}_t[dp_t]$ is a non-zero constant.

where $skew_t \equiv skew_t(p_{t+\Delta t})(\Delta t)^{-1/2}$.

The model has two predictions for the results of this regression: the coefficient on $\frac{1}{3}skew_t dp_t$ should be equal to 1, and the coefficient on $vol_t^2 dt$ should be equal to $-\sigma_Y^{-1}$. The first relationship 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 a structural parameter nonparametrically (and it will also be part of a cross-equation restriction). Additionally, we test the model's prediction that the mean reversion is quadratic rather than linear.

7.1.2 The κ_3 equation

Equation (10) gives the dynamics of the third moment of beliefs. Empirically, we can get the third moment either from an estimate of the conditional skewness of returns multiplied by volatility, via equation (17), or directly from volatility of volatility with equation (19). The fourth cumulant of beliefs appears on the right-hand side of equation (10). In principle that can be obtained from an estimate of the fourth moment of returns, but in practice we find that that method performs very poorly.²² The more effective way is to use an estimate of the skewness of volatility in equation (20).

Assuming $\mathbb{E}_t[dp_t]$ is zero, we then have two regression specifications from equation (10):

$$d[skew_t vol_t] = cons + \text{sign}(skew_t) skew_t vol_t \left(\frac{1}{3} skew_t \right) dp_t - 3\sigma_Y^{-1} skew_t vol_t^2 dt, \quad (28)$$

and

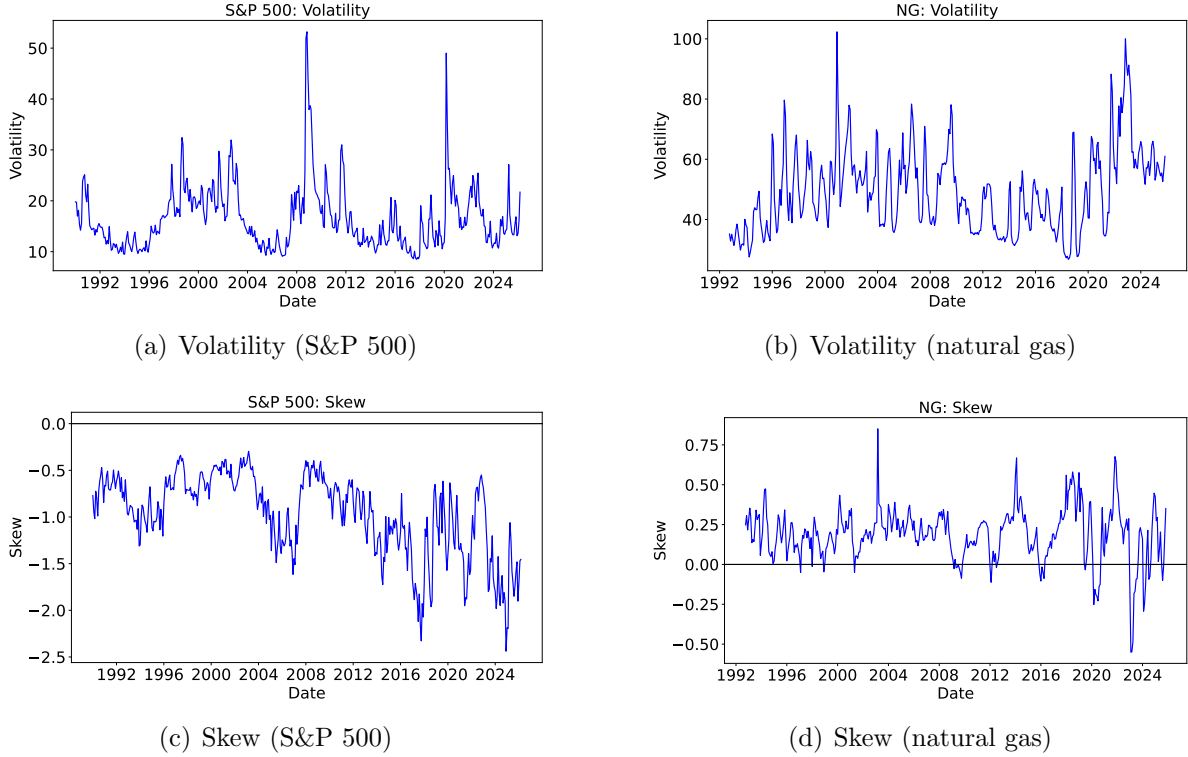
$$d[vvol_t] = cons + \text{sign}(skew_t) \frac{\frac{1}{3} skew_t vol_t vvol_t}{vol_t} dp_t - 3\sigma_Y^{-1} vol_t vvol_t dt, \quad (29)$$

where $skewvol_t \equiv skew_t(vol_{t+\Delta t})(\Delta t)^{-1/2}$ and $vvol_t$ is the instantaneous volatility of vol_t . The $\text{sign}(skew_t)$ factor arises because $vvol_t$ and $skewvol_t$ measure only the magnitude of the third moment, $|\kappa_{3,t}|$ (equations (19) and (20)), while the loading of the third moment on return surprises carries the sign of $\kappa_{3,t}$.

Under the model, if all of the objects can be estimated without error, the regressions should give identical results. In practice, estimating third moments tends to be more difficult than estimating second moments, which suggests that the specification using

²²Mathematically, the fourth moment of returns depends on both the third and fourth moments of beliefs, and the third moment contribution is dominant. Noise in the various estimators then makes getting the fourth moment of beliefs this way essentially infeasible, since it requires subtracting a function of the estimated third moment from the estimated fourth moment.

Figure 6: Time series of conditional volatility and skewness for S&P 500 and 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.

volatility of volatility should perform better.

7.2 Data

We estimate regressions for two markets. The first, given our focus on the stock market, is the S&P 500. For that case, we use conditional moments constructed as described in section 2. Similar to vol_t and $skew_t$, $vvol_t$ and $skewvol_t$ are constructed by forecasting the realized second and third moments of conditional volatility, where the option-based information, which is the main driver of the forecasts, comes from VIX options, which allow the calculation of the risk-neutral distribution of volatility.

Figure 6 plots the conditional volatility and skewness for the S&P 500.

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, predominantly positive skewness. To calculate

dp_t in this case we use natural gas futures returns from the CME. We then use options on futures to estimate the conditional moments in the same way as for the S&P 500. The time series of conditional volatility and skewness for natural gas are plotted in the right-hand panels of figure 6.

For natural gas, volatility derivatives are not exchange traded, so in that case we do not estimate the κ_3 regressions in equations (28) and (29) because there is insufficient data to get credible high-frequency conditional moments of volatility ($vvol_t$ and $skewvol_t$).

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.

For all the regressions, any time the same variable appears on both the left- and right-hand sides of the regressions (e.g. vol_{t-1} in (27)), we enter it with one additional lag to avoid any bias coming from measurement error. Similar results hold when using the second lag as an instrument for the first lag.

In constructing conditional moments, our baseline is to forecast realized moments using information from option prices. That is, the moments are conditional on linear functions of option-implied moments. That allows for some variation in risk premia, but only to the extent it can be spanned by the option moments themselves. As an alternative, we also project the realized moments on lagged realized moments, the Baa–Aaa credit spread, the one-month lagged CAPE from Shiller, and, for natural gas, the log futures-spot basis.

A final but critical point to note is that the results on moments in the model are all for a time horizon approaching zero, whereas in the data we follow the typical practice of using a one-month horizon. That naturally creates a potential wedge. To address that, below we replicate the empirical regressions in the quantitative simulations. The reason we report results constructing conditional moments only from option information is that that procedure can also be implemented in the simulations.

7.3 Results

7.3.1 κ_2 regressions

Table 3 reports results of the regression implied by (27). Theoretical values for the coefficients are listed on the left side of the table. 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 $\frac{1}{3}skew_t dp_t$ should be equal to 1, and in both cases that is a very good description of the data. Looking at the top panel, the coefficient is 1.03 with a standard error of 0.02 for the S&P 500 and 1.04 with a standard error of 0.05 for natural gas. The results are similar in the bottom panel using a richer set of predictors to construct the conditional moments.

That coefficient tests the hypothesis discussed in corollary 2 that skewness and the leverage effect coefficient should be proportional, since $\frac{1}{3}skew_t$ here is playing the role of the leverage effect coefficient. 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. The equality of that coefficient with 1 is therefore not a test of the model’s detailed learning specification but rather its assumption that prices and volatility are driven by the same diffusion.

A stricter test of the model is its implication that the volatility dynamics are quadratic. To evaluate that, the second and fourth columns of table 3 include both vol_{t-2}^2 and vol_{t-2} . The two regressors are highly collinear, so the individual coefficients are estimated much less precisely. For the stock market, the coefficient on vol_{t-2} is essentially zero, as the model predicts, while the point estimate on vol_{t-2}^2 remains negative and close to its value in column (1), and the fit is unchanged. For natural gas the quadratic term remains negative and statistically significant, though the linear term also enters with a small positive coefficient. The results are again highly similar in the top and bottom panels.

The estimated coefficients on vol_t^2 map to $-\sigma_Y^{-1}$ in the model. The confidence band for the stock market implies a value for σ_Y between 1.17 and 3.60, somewhat below (though of the same order of magnitude as) the value of 4.5 used in the simulations in the previous section. That wedge could be due to a gap caused by mismatch between the horizon of these regressions compared to the theoretical limit of zero time horizon (see section 7.4). For natural gas, the mean reversion of volatility appears much slower, which implies that σ_Y is larger and that investors have less precise signals about fundamentals.

7.3.2 κ_3 regression

Table 4 reports results of the regressions in equations (28) and (29). The rows are labeled as $\kappa_{4,t-1} dp_t$ and $3\kappa_{2,t-1}\kappa_{3,t-1} dt$ terms. They involve different regressors since the moments can be measured in different ways, but under the model the coefficients should be equal. The theoretical values are noted on the left side of the table. Additionally, the dt coefficient should be equal to the coefficient on vol_{t-2}^2 in table 3 – in both cases equal to $-\sigma_Y^{-1}$.

Column (1) reports results for Δvol_t on the left-hand side. The $\kappa_{4,t-1} dp_t$ coefficient

Table 3: Volatility regressions

		S&P 500		Natural gas	
Theoretical val.		(1)	(2)	(3)	(4)
Panel A: implied-moment predictors only					
Vol_{t-2}^2	$-\sigma_Y^{-1}$	−0.5670*** [0.1475]	−0.3993 [0.4802]	−0.1370*** [0.0401]	−0.6099** [0.2520]
$\frac{1}{3\sqrt{21}}skew_{t-1}dp_t$	1	1.0298*** [0.0248]	1.0296*** [0.0247]	1.0355*** [0.0531]	1.0342*** [0.0530]
Vol_{t-2}	0		−0.0056 [0.0117]		0.0346** [0.0163]
Observations		9126	9126	8303	8303
R^2		0.5891	0.5891	0.2164	0.2180
Panel B: additional predictors					
Vol_{t-2}^2	$-\sigma_Y^{-1}$	−0.3956* [0.2190]	−0.2734 [0.5252]	−0.1736*** [0.0450]	−0.6957** [0.2875]
$\frac{1}{3\sqrt{21}}skew_{t-1}dp_t$	1	1.0231*** [0.0342]	1.0229*** [0.0342]	0.8987*** [0.0480]	0.8975*** [0.0479]
Vol_{t-2}	0		−0.0055 [0.0148]		0.0370** [0.0181]
Observations		9064	9064	7249	7249
R^2		0.4443	0.4444	0.1992	0.2011

Note: Daily regressions of first differences in volatility (for S&P 500 and natural gas) onto different predictors. In panel A the risk-adjusted moments are constructed by projecting realized moments on option-implied moments only; in panel B each projection additionally conditions on predetermined predictors: its own trailing 63-day realized moment, the Baa–Aaa credit spread, one-month-lagged CAPE and, for natural gas, the log futures–spot basis (which limits that sample to 1997 forward, when the Henry Hub spot series begins). Stars indicate statistical significance: * $p < .1$, ** $p < .05$, *** $p < .01$.

is, as predicted, reasonably close to -1 . The $3\kappa_{2,t-1}\kappa_{3,t-1}dt$ coefficient is -0.39 in the top panel and -0.25 in the bottom panel. That is smaller than the value in table 3, but the hypothesis that they are equal is not rejected in either case. So not only does the model’s prediction for the form of nonlinearity in the dynamics appear to describe the data well, but also the magnitudes are linked across equations.

Column (2) reports results from the regression based on equation (28). In this case, the $3\kappa_{2,t-1}\kappa_{3,t-1}dt$ coefficient is again consistent with the one in table 3 and it is also consistent with column (1) in table 4. On the other hand, the coefficient on $\kappa_{4,t-1}dp_t$, which should be equal to -1 , is instead close to zero. One view of this is that it is a rejection of the model. At the same time, conditional skewness is the most difficult moment to estimate

Table 4: Third-moment regressions

		Δvol_t	$\Delta \left[\frac{1}{3\sqrt{21}} vol_t skew_t \right]$
	Theoretical val.	(1)	(2)
Panel A: implied-moment predictors only			
$\kappa_{4,t-1} dp_t$	-1	-0.7495*** [0.0682]	-0.1240*** [0.0335]
$3\kappa_{2,t-1}\kappa_{3,t-1} dt$	$-\sigma_Y^{-1}$	-0.3895*** [0.1394]	-0.4213*** [0.1618]
p -value: dt coef. = Vol_{t-2}^2 (table 3)		0.101	0.528
Observations		4885	4912
R^2		0.4213	0.0190
Panel B: additional predictors			
$\kappa_{4,t-1} dp_t$	-1	-1.0474*** [0.0436]	0.1826*** [0.0220]
$3\kappa_{2,t-1}\kappa_{3,t-1} dt$	$-\sigma_Y^{-1}$	-0.2464** [0.0997]	-0.2530* [0.1313]
p -value: dt coef. = Vol_{t-2}^2 (table 3)		0.278	0.601
Observations		4903	4903
R^2		0.4357	0.0425

Note: Daily S&P 500 regressions of the model's two representations of the third-moment dynamics: column (1) has the change in the volatility of volatility on the left-hand side (equation (29)); column (2) the change in $\frac{1}{3\sqrt{21}} vol_t skew_t$ (equation (28)). Constants are estimated but not reported. Panels A and B are constructed as in table 3. The return terms use the risk-adjusted skewness of volatility, constructed from VIX options like the other conditional moments, smoothed over 63 days. Stars indicate statistical significance: * $p < .1$, ** $p < .05$, *** $p < .01$.

in the data, both because it is sensitive to the prices of deeply out of the money options and also because realized skewness, being a higher power, is much noisier than realized volatility. Equation (28) has three more instances of conditional skewness than equation (29), and the coefficient that is most problematic involves a product of two different conditional skews.

When estimating σ_Y from the coefficients on $3\kappa_{2,t-1}\kappa_{3,t-1} dt$ as was done with the results in table 3, the confidence bands for σ_Y implied by the regressions in table 4 are [1.51, 8.60] and [1.35, 9.60], which both contain the value of 4.5 used in the simulation.

Overall, the model's predictions hold both within and across the three regressions with the exception of the $\kappa_{4,t} dp_t$ coefficient measured from an interaction of conditional skewness terms. While, as noted above, the κ_2 regression results by themselves hold as long as volatility and returns are well described as being jointly driven by a single diffusion,

the link in the coefficients across the κ_2 and κ_3 regressions is a tighter prediction of the model.

7.4 Comparing the empirical regressions to the simulated model

As noted above, one wedge between the regressions and the theoretical predictions from the model is that the empirical regressions are based on conditional moments for returns over the next month (then properly rescaled), whereas the predictions are limits as the time horizon goes to zero. It is straightforward, though, to implement exactly the same procedure used to generate the regression results in the simulations of the model. Specifically, that can be done for the version of the regressions where the predictors for the realized moments are the option-implied moments, since those can be calculated in the simulation.

Table 5 reports regressions using identical methods in the data and the model. The numbers in brackets for the model columns are 95% sampling intervals, where each simulated sample is the same length as the empirical data. For the κ_2 regressions, the data estimate for the coefficient on lagged vol_t^2 is nearly identical to the average from the simulations. The coefficient on $skew_{t-1}dp_t$ in the data is very close to 1, but larger in the simulations. That same phenomenon occurs for the κ_3 regression in the bottom panel of the table. Notably, that regressor is the one with skewness in it, and so it is possible that measurement error attenuates the coefficient.

Finally, the drift coefficient in the bottom panel is, similar to the top panel, very close to the value from the simulations. The two drift (dt) coefficients are theoretically related to σ_Y , since they determine the rate of mean reversion in the moments. The fact that the values from the simulations are consistent with what is observed empirically validates the calibrated choice of σ_Y .

7.5 Estimates of investors' uncertainty about fundamentals

The estimated regression coefficients on vol_{t-1}^2 give an estimate of $-\sigma_Y^{-1}$ at the daily level (given that these are daily regressions), which can then be combined with the volatility and skewness of stock market returns to reveal the standard deviation and skewness of agents' posteriors for fundamentals. Specifically, recall that when $a_t = 0$,

$$vol_t = \kappa_{2,t}\sigma_Y^{-1} \quad \Rightarrow \quad \kappa_{2,t}^{1/2} = (vol_t\sigma_Y)^{1/2} \quad (30)$$

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

Table 5: Volatility and third-moment regressions: data vs. model

Panel A: volatility regression (κ_2 equation)

	Theoretical val.	Δvol_t	
		Data	Model
Vol_{t-2}^2	$-\sigma_Y^{-1}$	-0.5670	-0.5173 [-0.9342, -0.2866]
$\frac{1}{3\sqrt{21}}skew_{t-1}dp_t$	1	1.0298	1.4791 [1.1752, 2.0232]
Observations		9126	9126
R^2		0.5891	0.8500

Panel B: third-moment regressions (κ_3 equation)

	Theoretical val.	$\Delta vvol_t$		$\Delta \left[\frac{1}{3\sqrt{21}}vol_tskew_t \right]$	
		Data	Model	Data	Model
$\kappa_{4,t-1}dp_t$	-1	-0.7495	-1.0809 [-3.0235, -0.3944]	-0.1240	-0.9141 [-2.6853, -0.1552]
$3\kappa_{2,t-1}\kappa_{3,t-1}dt$	$-\sigma_Y^{-1}$	-0.3895	-0.4693 [-1.0194, -0.1305]	-0.4213	-0.7176 [-1.5755, -0.2050]
Observations		4885	4885	4912	4912
R^2		0.4213	0.6797	0.0190	0.5264

Note: Each panel pairs the empirical point estimates (“Data”; standard errors and significance levels are reported in tables 3 and 4) with the identical procedure applied to the model’s simulated path (“Model”). Brackets are the 2.5th and 97.5th percentiles of the simulated coefficient distribution. The top panel repeats the S&P 500 volatility regression of column (1) of table 3; the bottom panel the third-moment regressions of table 4 (both from the implied-moment-predictors-only specification, panel A of those tables). Constants are estimated but not reported.

σ_Y is between 1.17 and 3.60, based on confidence band for the coefficient on vol_{t-1}^2 in equation (27). Taking equation (30) and inserting the value for σ_Y along with the historical daily standard deviation of stock returns in our sample, 1.14 percent, implies that agents’ posterior standard deviation is between 11.5 and 20.3 percent. The ± 2 standard deviation range for fundamentals around the current price for the aggregate stock market is then between ± 23.1 and ± 40.5 percent.

Similarly, we can get an estimate of average skewness in beliefs. One-month conditional return skewness is historically approximately -1 . Plugging that into (17) along with the estimates of κ_2 and σ_Y yields an estimate for the skewness of fundamentals between -0.74 and -1.29 . 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 .

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. While many of the results here have been shown to hold in specific learning settings, the contribution of this paper is to show that they are generic predictions and to give necessary and sufficient conditions for the model to match features of the data like the leverage effect.

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.

References

- 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, Rüdiger, Kai Carstensen, Stefan Lautenbacher, 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* 2009.
- 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.
- Bekaert, Geert and Guojun Wu, “Asymmetric volatility and risk in equity markets,”

- The review of financial studies*, 2000, 13 (1), 1–42.
- Bianchi, Francesco, Cosmin Ilut, and Hikaru Saijo**, “Smooth Diagnostic Expectations,” 2025. Working paper.
- 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.
- 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, 88, 185–207.
- Campbell, John Y. and Ludger Hentschel**, “No news is good news: An asymmetric model of changing volatility in stock returns,” *Journal of Financial Economics*, 1992, 31 (3), 281–318.
- Christiano, Lawrence J., Martin Eichenbaum, and Benjamin K. Johansson**, “Slow Learning,” 2026. Working paper.
- 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.
- David, Alexander 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.
- 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.
- Eusepi, Stefano** and **Bruce Preston**, “Expectations, learning, and business cycle fluctuations,” *American Economic Review*, 2011, *101* (6), 2844–2872.
- Farmer, Leland E**, **Emi Nakamura**, and **Jón Steinsson**, “Learning about the long run,” *Journal of Political Economy*, 2024, *132* (10).
- 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.
- Ghaderi, Mohammad**, **Mete Kilic**, and **Sang Byung Seo**, “Learning, slowly unfolding disasters, and asset prices,” *Journal of Financial Economics*, 2022, *143* (1), 527–549.
- 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.
- Ilut, Cosmin** and **Hikaru Saijo**, “Learning, confidence, and business cycles,” *Journal of Monetary Economics*, 2021, *117*, 354–376.
- Iseringhausen, Martin**, **Ivan Petrella**, and **Konstantinos Theodoridis**, “Aggregate skewness and the business cycle,” *Review of Economics and Statistics*, 2023, pp. 1–38.
- Kozlowski, Julian**, **Laura Veldkamp**, and **Venky Venkateswaran**, “The tail that keeps the riskless rate low,” *NBER Macroeconomics Annual*, 2018, *33* (1), 253–283.
- , —, and —, “The Tail That Wags the Economy: Beliefs and Persistent Stagnation,” *Journal of Political Economy*, 2020, *128* (8).
- Liptser, Robert S** and **Albert N Shiryaev**, *Statistics of random processes: I. General theory*, Springer Science & Business Media, 2013.
- Lorenzoni, Guido**, “A theory of demand shocks,” *American economic review*, 2009, *99* (5), 2050–2084.
- 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*.
- 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.
- 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.
- Molavi, Pooya**, “The Empirical Content of Bayesian Updating under Misspecification,” 2026. Working paper.
- 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.
- Nimark, Kristoffer P**, “Man-bites-dog business cycles,” *American Economic Review*, 2014, 104 (8), 2320–2367.
- Orlik, Anna and Laura Veldkamp**, “Understanding uncertainty shocks and the role of black swans,” *Journal of Economic Theory*, 2024, p. 105905.
- Osband, Kent**, “Rational Turbulence,” 2011. Working paper.
- , “Rationally Turbulent Expectations Chapter 4: Cumulant Hierarchy,” 2025. SSRN working paper.
- Protter, Philip E.**, *Stochastic Integration and Differential Equations*, Springer, 2005.
- Rietz, Thomas A.**, “The Equity Risk Premium: A Solution,” *Journal of Monetary Economics*, 1988, 22(1), 117–131.
- Roll, Richard**, “ R^2 ,” *Journal of Finance*, 1988.
- 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.

Taleb, Nassim Nicholas and Raphael Douady, “Mathematical definition, mapping, and detection of (anti) fragility,” *Quantitative Finance*, 2013, 13 (11), 1677–1689.

van Nieuwerburgh, Stijn and Laura Veldkamp, “Information acquisition and under-diversification,” *The Review of Economic Studies*, 2010, 77 (2), 779–805.

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 Proofs

A.1 Proof of theorem 1

A.1.1 Assumptions

Assumption 3 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 $\int_0^t \mathbb{E}[|\mathcal{G}_{\omega,s}|^2]ds < \infty$ for all finite t such that $\phi_{\omega,t} - \phi_{\omega,0} - \int_0^t \mathcal{G}_{\omega,s}ds$ is a right-continuous square-integrable martingale. Moreover, the map $\omega \mapsto \mathcal{G}_{\omega,t}$ is n times differentiable in a neighborhood of zero with its derivatives up to order n locally dominated by integrable processes.

Assumption 4 $\int_0^t \mathbb{E}[x_s^2]ds < \infty$ for all finite t .

Assumption 5 Let $\langle W, \exp(i\omega x) \rangle$ denote the predictable quadratic covariation of W and the process $\exp(i\omega x)$. The map $\omega \mapsto d\langle W, \exp(i\omega x) \rangle_t/dt$ is n times differentiable in a neighborhood of zero with its first n derivatives dominated by conditionally integrable random variables.

A.1.2 Intermediate results

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 3 and 4 are satisfied, then*

$$d\varphi_{x,t}(\omega) = \mathbb{E}_t[d\exp(i\omega x_t)] + \left(\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right] + \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\sigma_Y} \right) \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_Y}, \quad (\text{A.1})$$

where $\mathbb{E}_t$ and cov_t denote the expectation and covariance operators, respectively, conditional on Y^t , and $\langle W, \exp(i\omega x) \rangle$ denotes the predictable quadratic covariation of W and the process $\exp(i\omega x)$.

The lemma follows from theorem 8.1 of [Liptser and Shiryaev \(2013\)](#) by setting $h_t \rightarrow \phi_{\omega,t}$, $\xi_t \rightarrow Y_t$, $A_t \rightarrow x_t$, and $B_t(\xi) \rightarrow \sigma_Y$. 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 3 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 4. The second part of condition (8.3) and conditions (8.4), (8.5), (8.9) are satisfied since σ_Y is time invariant. Condition (8.6) is satisfied since $\phi_{\omega,t}$ is a bounded function. Finally, theorem 8.1 gives, in differential form,

$$d\mathbb{E}_t[\exp(i\omega x_t)] = \mathbb{E}_t[\mathcal{G}_{\omega,t}]dt + \left(\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right] + \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\sigma_Y} \right) \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_Y}. \quad (\text{A.2})$$

On the other hand, by the definition of $\mathcal{G}_{\omega,t}$, $d\exp(i\omega x_t) - \mathcal{G}_{\omega,t}dt = dM_{\omega,t}$, where $M_{\omega,t}$ is a right-continuous martingale. Therefore, $\mathbb{E}_t[\mathcal{G}_{\omega,t}]dt = \mathbb{E}_t[d\exp(i\omega x_t)]$.

Lemma 2 *Suppose assumptions 3–5 are satisfied. Let $\kappa_{k,t}$ denote the k th cumulant of the posterior distribution of x_t conditional on Y^t . Suppose the $(n+1)$ th posterior cumulant exists and $\mathbb{E}_t[d(x_t^j)]$ exists for each $j \leq n$. Then for every $k \leq n$,*

$$\begin{aligned} d\kappa_{k,t} = & \sum_{j=1}^k \binom{k}{j} B_{k-j}(-\kappa_{1,t}, \dots, -\kappa_{k-j,t}) \mathbb{E}_t[d(x_t^j)] + \frac{\tilde{\kappa}_{k+1,t}}{\sigma_Y^2} (dY_t - \mathbb{E}_t[x_t]dt) \\ & - \frac{1}{2\sigma_Y^2} \sum_{j=2}^k \binom{k}{j-1} \tilde{\kappa}_{j,t} \tilde{\kappa}_{k-j+2,t} dt, \end{aligned} \quad (\text{A.3})$$

where

$$\tilde{\kappa}_{k,t} = \kappa_{k,t} + \sigma_Y \sum_{j=1}^{k-1} \binom{k-1}{j} B_{k-1-j}(-\kappa_{1,t}, \dots, -\kappa_{k-1-j,t}) \mathbb{E}_t \left[\frac{d\langle W, x^j \rangle_t}{dt} \right], \quad (\text{A.4})$$

for $2 \leq k \leq n+1$ and B_j denotes the j th complete exponential Bell polynomial.

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

$$\begin{aligned} d \log \varphi_{x,t}(\omega) &= \frac{\mathbb{E}_t[d \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} + \frac{\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right]}{\mathbb{E}_t[\exp(i\omega x_t)]} \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_Y} \\ &\quad + \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} \frac{dY_t - \mathbb{E}_t[x_t]dt}{\sigma_Y^2} \\ &\quad - \frac{1}{2} \left(\frac{\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right]}{\mathbb{E}_t[\exp(i\omega x_t)]} + \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\sigma_Y \mathbb{E}_t[\exp(i\omega x_t)]} \right)^2 dt \end{aligned} \quad (\text{A.5})$$

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

$$\kappa_{k,t} = i^{-k} \frac{d^k}{d\omega^k} \log \varphi_{x,t}(\omega) \Big|_{\omega=0} \quad (\text{A.6})$$

for any $k \leq n+1$ (see chapter 2 of [Lukacs \(1970\)](#) for the results on characteristic functions and cumulants used here). Taking the Itô differential of the above equation (and applying the dominated convergence theorem to switch the order of d and $d^k/d\omega^k$),

$$d\kappa_{k,t} = i^{-k} \frac{d^k}{d\omega^k} (d \log \varphi_{x,t}(\omega)) \Big|_{\omega=0}. \quad (\text{A.7})$$

The remainder of the proof calculates the k th derivative of the right-hand side of (A.5) for $k \leq n$.

For the **first term**, the second part of assumption 3 implies that, for any ω in a sufficiently small neighborhood of the origin,

$$\mathbb{E}_t[d \exp(i\omega x_t)] = \sum_{j=0}^n \frac{(i\omega)^j}{j!} \mathbb{E}_t[d(x_t^j)] + o(\omega^n). \quad (\text{A.8})$$

Therefore,

$$\frac{d^k}{d\omega^k} \mathbb{E}_t[d \exp(i\omega x_t)] \Big|_{\omega=0} = i^k \sum_{j=k}^n \frac{(i\omega)^{j-k}}{(j-k)!} \mathbb{E}_t[d(x_t^j)] \Big|_{\omega=0} = i^k \mathbb{E}_t[d(x_t^k)]. \quad (\text{A.9})$$

The Leibniz rule then yields

$$\frac{d^k}{d\omega^k} \frac{\mathbb{E}_t[d \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} \Big|_{\omega=0} = \sum_{j=0}^k \binom{k}{j} \left(\frac{d^{k-j}}{d\omega^{k-j}} \mathbb{E}_t[d \exp(i\omega x_t)] \Big|_{\omega=0} \right) \left(\frac{d^j}{d\omega^j} (\mathbb{E}_t[\exp(i\omega x_t)])^{-1} \Big|_{\omega=0} \right). \quad (\text{A.10})$$

Finally, note that $(\mathbb{E}_t[\exp(i\omega x_t)])^{-1} = \exp(-\log \varphi_{x,t}(\omega))$ and that the complete exponential

Bell polynomials can be used to transform the right-hand side of (A.10) into (see Comtet (1974) section 3.3)

$$\frac{d^k}{d\omega^k} \frac{\mathbb{E}_t[d \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} \Big|_{\omega=0} = i^k \sum_{j=1}^k \binom{k}{j} B_{k-j}(-\kappa_{1,t}, \dots, -\kappa_{k-j,t}) \mathbb{E}_t[d(x_t^j)]. \quad (\text{A.11})$$

For the **second term** on the right-hand side of (A.5), the differentiability and domination assumptions on $d\langle W, \exp(i\omega x) \rangle_t / dt$ imply that, for any ω in a sufficiently small neighborhood of the origin,

$$\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right] = \sum_{j=0}^n \frac{(i\omega)^j}{j!} \mathbb{E}_t \left[\frac{d\langle W, x^j \rangle_t}{dt} \right] + o(\omega^n). \quad (\text{A.12})$$

Since $d\langle W, x^0 \rangle_t = 0$,

$$\frac{d^k}{d\omega^k} \mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right] \Big|_{\omega=0} = i^k \mathbb{E}_t \left[\frac{d\langle W, x^k \rangle_t}{dt} \right]. \quad (\text{A.13})$$

The same argument as for the first term then yields

$$\frac{d^k}{d\omega^k} \frac{\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right]}{\mathbb{E}_t[\exp(i\omega x_t)]} \Big|_{\omega=0} = i^k \sum_{j=1}^k \binom{k}{j} B_{k-j}(-\kappa_{1,t}, \dots, -\kappa_{k-j,t}) \mathbb{E}_t \left[\frac{d\langle W, x^j \rangle_t}{dt} \right]. \quad (\text{A.14})$$

Denoting the sum on the right-hand side by $\Lambda_{k,t}$, the k th derivative of the second term in (A.5), evaluated at $\omega = 0$, is $i^k \Lambda_{k,t}(dY_t - \mathbb{E}_t[x_t]dt)/\sigma_Y$.

For the **third term** on the right-hand side of (A.5), we have

$$\frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} = \frac{\mathbb{E}_t[x_t \exp(i\omega x_t)]}{\mathbb{E}_t[\exp(i\omega x_t)]} - \mathbb{E}_t[x_t] = i^{-1} \frac{d}{d\omega} \log \varphi_{x,t}(\omega) - \mathbb{E}_t[x_t]. \quad (\text{A.15})$$

Therefore,

$$\frac{d^k}{d\omega^k} \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\mathbb{E}_t[\exp(i\omega x_t)]} \Big|_{\omega=0} = i^{-1} \frac{d^{k+1}}{d\omega^{k+1}} \log \varphi_{x,t}(\omega) \Big|_{\omega=0} = i^k \kappa_{k+1,t}, \quad (\text{A.16})$$

and the k th derivative of the third term in (A.5), evaluated at $\omega = 0$, is $i^k \kappa_{k+1,t}(dY_t - \mathbb{E}_t[x_t]dt)/\sigma_Y^2$. The derivatives of the second and third terms combine to $i^k \tilde{\kappa}_{k+1,t}(dY_t - \mathbb{E}_t[x_t]dt)/\sigma_Y^2$, where $\tilde{\kappa}_{k,t} = \kappa_{k,t} + \sigma_Y \Lambda_{k-1,t}$.

Finally, for the **fourth term** in (A.5), the Leibniz rule combined with the results above gives

$$\frac{d^k}{d\omega^k} \left(\frac{\mathbb{E}_t \left[\frac{d\langle W, \exp(i\omega x) \rangle_t}{dt} \right]}{\mathbb{E}_t[\exp(i\omega x_t)]} + \frac{\text{cov}_t(x_t, \exp(i\omega x_t))}{\sigma_Y \mathbb{E}_t[\exp(i\omega x_t)]} \right)^2 \Big|_{\omega=0} \quad (\text{A.17})$$

$$= i^k \sum_{r=1}^{k-1} \binom{k}{r} \left(\Lambda_{r,t} + \frac{\kappa_{r+1,t}}{\sigma_Y} \right) \left(\Lambda_{k-r,t} + \frac{\kappa_{k-r+1,t}}{\sigma_Y} \right) \quad (\text{A.18})$$

$$= i^k \sum_{j=2}^k \binom{k}{j-1} \frac{\tilde{\kappa}_{j,t} \tilde{\kappa}_{k-j+2,t}}{\sigma_Y^2}. \quad (\text{A.19})$$

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

A.1.3 Proof

Assumption 3 implies that x is a semimartingale with no fixed-time jumps. By Itô's lemma for semimartingales (Protter, 2005, p. 79),

$$x_t^j = x_0^j + j \int_{0+}^t x_{s-}^{j-1} dx_s + A_t, \quad (\text{A.20})$$

where A_t is a finite variation process. By bracket rule for stochastic integrals (Protter, 2005, p. 75), for any $j \geq 1$,

$$\frac{d\langle W, x^j \rangle_t}{dt} = j x_t^{j-1} \alpha_t. \quad (\text{A.21})$$

Since x has no fixed-time jumps and α_t is Y^t -measurable,

$$\mathbb{E}_t \left[\frac{d\langle W, x^j \rangle_t}{dt} \right] = j \alpha_t \mathbb{E}_t[x_t^{j-1}]. \quad (\text{A.22})$$

Therefore, using $\mathbb{E}_t[x_t^j] = B_j(\kappa_{1,t}, \dots, \kappa_{j,t})$, for all $3 \leq k \leq n+1$,

$$\sum_{j=1}^{k-1} \binom{k-1}{j} B_{k-1-j}(-\kappa_{1,t}, \dots, -\kappa_{k-1-j,t}) \mathbb{E}_t \left[\frac{d\langle W, x^j \rangle_t}{dt} \right] \quad (\text{A.23})$$

$$= \alpha_t (k-1) \sum_{j=0}^{k-2} \binom{k-2}{j} B_{k-2-j}(-\kappa_{1,t}, \dots, -\kappa_{k-2-j,t}) B_j(\kappa_{1,t}, \dots, \kappa_{j,t}) \quad (\text{A.24})$$

$$= \alpha_t (k-1) B_{k-2}(0, \dots, 0) = 0. \quad (\text{A.25})$$

Therefore, $\tilde{\kappa}_{k,t} = \kappa_{k,t}$ for all $3 \leq k \leq n+1$. On the other hand,

$$\tilde{\kappa}_{2,t} = \kappa_{2,t} + \sigma_Y B_0 \mathbb{E}_t \left[\frac{d\langle W, x \rangle_t}{dt} \right] = \kappa_{2,t} + \sigma_Y \alpha_t. \quad (\text{A.26})$$

Substituting these expressions for $\tilde{\kappa}_{k,t}$ into lemma 2 yields equation (11).

A.2 Proof of proposition 2

The dynamics of the first moment are as in theorem 1 with α_t set equal to $a_t \sigma_Y^{-1} \kappa_{2,t}$:

$$dp_t = \kappa_{2,t} (1 + a_t) \sigma_Y^{-2} (dY_t - \mathbb{E}_t[x_t] dt) + \mathbb{E}_t[dx_t]. \quad (\text{A.27})$$

Taking the conditional expectation of the above expression yields $\mathbb{E}_t[dp_t] = \mathbb{E}_t[dx_t]$. The expression in the proposition follows by substituting for $\mathbb{E}_t[dx_t]$ with $\mathbb{E}_t[dp_t]$.

For the second moment, we simply need to show that

$$\sum_{j=1}^2 \binom{2}{j} B_{2-j}(-\kappa_{1,t}, \dots, -\kappa_{2-j,t}) \mathbb{E}_t[d(x_t^j)] = \text{var}_t(dx_t) + 2 \text{cov}_t(x_{t-}, dx_t). \quad (\text{A.28})$$

The left-hand side of the above expression equals $\mathbb{E}_t[d(x_t^2)] - 2\mathbb{E}_t[x_t]\mathbb{E}_t[dx_t]$. Assumption 3 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)),

$$\mathbb{E}_t[d(x_t^2)] = 2\mathbb{E}_t[x_{t-}dx_t] + \text{var}_t(dx_t), \quad (\text{A.29})$$

where $x_{t-} \equiv \lim_{s \uparrow t} x_s$ and $\text{var}_t(dx_t)$ is the shorthand for $\mathbb{E}_t[d[x]_t]$. By assumption 3, $x_t = x_{t-} + \Delta M_t$, where ΔM_t denotes the time- t jump of M . Since Y_t is a continuous process, its natural filtration is continuous. Therefore,

$$\mathbb{E}_t[\Delta M_t] = \mathbb{E}_t[\Delta M_t | Y^t] = \mathbb{E}_t[\Delta M_t | Y^{t-}] = 0, \quad (\text{A.30})$$

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

$$\mathbb{E}_t[d(x_t^2)] - 2\mathbb{E}_t[x_t]\mathbb{E}_t[dx_t] = \text{var}_t(dx_t) + 2 \text{cov}_t(x_{t-}, dx_t), \quad (\text{A.31})$$

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

Finally, the dynamics of the third moment are as in theorem 1 with α_t set equal to $a_t \sigma_Y^{-1} \kappa_{2,t}$.

A.3 Proof of corollary 1

Since a_t is absolutely continuous, it is a finite variation process and therefore $d[a, \kappa_2]_t = 0$, where $[a, \kappa_2]$ denotes the quadratic covariation of a and κ_2 . Using equation (13) and the product rule,

$$d(vol_t) = (1 + a_t) \sigma_Y^{-1} d\kappa_{2,t} + \sigma_Y^{-1} \kappa_{2,t} da_t. \quad (\text{A.32})$$

Equation (8) implies

$$dp_t - \mathbb{E}_t[dp_t] = \kappa_{2,t} (1 + a_t) \sigma_Y^{-2} (dY_t - \mathbb{E}_t[x_t]dt). \quad (\text{A.33})$$

Substituting equation (9) into the first expression and using the above equation together with $vol_t^2 = \kappa_{2,t}^2 (1 + a_t)^2 \sigma_Y^{-2}$ gives equation (14).

A.4 Proof of proposition 3

Since a_t is absolutely continuous, all terms in equation (14) other than the first are finite variation and therefore do not contribute to the instantaneous covariance with dp_t . Hence,

$$\text{cov}_t(dp_t, d\text{vol}_t) = \sigma_Y^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}} \text{var}_t(dp_t). \quad (\text{A.34})$$

Dividing both sides by $\text{var}_t(dp_t)$ gives equation (15).

A.5 Proof of proposition 4

Let $d\bar{W}_t = (dY_t - \mathbb{E}_t[x_t]dt) / \sigma_Y$ denote the innovation process. The second order Itô-Taylor expansion of the price process using proposition 2 gives

$$\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2] = (1 + a_t)^2 \kappa_{2,t}^2 \sigma_Y^{-2} \Delta t + o(\Delta t), \quad (\text{A.35})$$

and

$$\begin{aligned} & \mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3] \\ &= \mathbb{E}_t \left[\left((1 + a_t) \kappa_{2,t} \sigma_Y^{-1} \Delta \bar{W}_t + \frac{(1 + a_t) \kappa_{3,t} \sigma_Y^{-2}}{2} (\Delta \bar{W}_t^2 - \Delta t) \right)^3 \right] + o(\Delta t^2) \end{aligned} \quad (\text{A.36})$$

$$= 3(1 + a_t)^3 \kappa_{2,t}^2 \kappa_{3,t} \sigma_Y^{-4} \Delta t^2 + o(\Delta t^2). \quad (\text{A.37})$$

Therefore,

$$\text{skew}_t(p_{t+\Delta t}) (\Delta t)^{-1/2} = \frac{\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^3] (\Delta t)^{-1/2}}{(\mathbb{E}_t[(p_{t+\Delta t} - \mathbb{E}_t[p_{t+\Delta t}])^2])^{3/2}} = 3 \frac{\kappa_{3,t}}{\kappa_{2,t}} \sigma_Y^{-1} + o(1). \quad (\text{A.38})$$

A.6 Proof of proposition 5

Using $\text{vol}_t = \kappa_{2,t} (1 + a_t) \sigma_Y^{-1}$ and applying the same second-order expansion to the volatility process,

$$\mathbb{E}_t[(\text{vol}_{t+\Delta t} - \mathbb{E}_t[\text{vol}_{t+\Delta t}])^2] = (1 + a_t)^2 \kappa_{3,t}^2 \sigma_Y^{-4} \Delta t + o(\Delta t), \quad (\text{A.39})$$

$$\mathbb{E}_t[(\text{vol}_{t+\Delta t} - \mathbb{E}_t[\text{vol}_{t+\Delta t}])^3] = 3(1 + a_t)^3 \kappa_{3,t}^2 \kappa_{4,t} \sigma_Y^{-7} \Delta t^2 + o(\Delta t^2). \quad (\text{A.40})$$

Therefore, for $\kappa_{3,t} \neq 0$,

$$\text{skew}_t(\text{vol}_{t+\Delta t}) (\Delta t)^{-1/2} = \frac{\mathbb{E}_t[(\text{vol}_{t+\Delta t} - \mathbb{E}_t[\text{vol}_{t+\Delta t}])^3] (\Delta t)^{-1/2}}{(\mathbb{E}_t[(\text{vol}_{t+\Delta t} - \mathbb{E}_t[\text{vol}_{t+\Delta t}])^2])^{3/2}} = 3 \frac{\kappa_{4,t}}{|\kappa_{3,t}|} \sigma_Y^{-1} + o(1). \quad (\text{A.41})$$

B Additional derivations

B.1 Equation (27)

Starting from equation (14) and setting $a_t = 0$,

$$d[vol_t] = \sigma_Y^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}} (dp_t - \mathbb{E}_t[dp_t]) - \sigma_Y^{-1} vol_t^2 dt + \sigma_Y^{-1} (\text{var}_t(dx_t) + 2 \text{cov}_t(x_{t-}, dx_t)). \quad (\text{B.1})$$

Equation (17) implies that

$$\sigma_Y^{-1} \frac{\kappa_{3,t}}{\kappa_{2,t}} = \frac{1}{3} skew_t. \quad (\text{B.2})$$

Therefore,

$$d[vol_t] = \frac{1}{3} skew_t (dp_t - \mathbb{E}_t[dp_t]) - \sigma_Y^{-1} vol_t^2 dt + \sigma_Y^{-1} (\text{var}_t(dx_t) + 2 \text{cov}_t(x_{t-}, dx_t)). \quad (\text{B.3})$$

If x has independent increments, then $\text{cov}_t(x_{t-}, dx_t) = 0$. If its increments are also stationary, then $\text{var}_t(dx_t)$ is a constant.

B.2 Equations (28) and (29)

By Itô's lemma for semimartingales (e.g. theorem 32 of Protter (2005))

$$d(x_t^2) = 2x_{t-} dx_t + d[x]_t, \quad (\text{B.4})$$

and

$$d(x_t^3) = 3x_{t-}^2 dx_t + 3x_{t-} d[x]_t + d\left(\sum_{0 < s \leq t} (\Delta x_s)^3\right), \quad (\text{B.5})$$

where Δx_s denotes the time- s jump of x . Since x has independent increments,

$$\mathbb{E}_t[d(x_t^2)] = 2\mathbb{E}_t[x_{t-}] \mathbb{E}_t[dx_t] + \mathbb{E}_t[d[x]_t], \quad (\text{B.6})$$

and

$$\mathbb{E}_t[d(x_t^3)] = 3\mathbb{E}_t[x_{t-}^2] \mathbb{E}_t[dx_t] + 3\mathbb{E}_t[x_{t-}] \mathbb{E}_t[d[x]_t] + \mathbb{E}_t\left[d\left(\sum_{0 < s \leq t} (\Delta x_s)^3\right)\right]. \quad (\text{B.7})$$

Assumption 3 implies that x has no fixed-time jumps. Therefore, $\mathbb{E}_t[x_{t-}] = \mathbb{E}_t[x_t] = \kappa_{1,t}$ and $\mathbb{E}_t[x_{t-}^2] = \mathbb{E}_t[x_t^2] = \kappa_{1,t}^2 + \kappa_{2,t}$. It follows that

$$\begin{aligned} & 3(\mathbb{E}_t[x_t]^2 - \kappa_{2,t}) \mathbb{E}_t[dx_t] - 3\mathbb{E}_t[x_t] \mathbb{E}_t[d(x_t^2)] + \mathbb{E}_t[d(x_t^3)] \\ &= 3(\kappa_{1,t}^2 - \kappa_{2,t}) \mathbb{E}_t[dx_t] - 3\kappa_{1,t} (2\kappa_{1,t} \mathbb{E}_t[dx_t] + \mathbb{E}_t[d[x]_t]) \\ &+ 3(\kappa_{1,t}^2 + \kappa_{2,t}) \mathbb{E}_t[dx_t] + 3\kappa_{1,t} \mathbb{E}_t[d[x]_t] + \mathbb{E}_t\left[d\left(\sum_{0 < s \leq t} (\Delta x_s)^3\right)\right] \end{aligned} \quad (\text{B.8})$$

$$= \mathbb{E}_t \left[d \left(\sum_{0 < s \leq t} (\Delta x_s)^3 \right) \right]. \quad (\text{B.9})$$

The final expression is a constant multiple of dt because the increments of x are stationary.

Equation (10) can then be written as

$$d\kappa_{3,t} = cons + \kappa_{4,t}\sigma_Y^{-2} (dY_t - \mathbb{E}_t[x_t]dt) - 3\sigma_Y^{-2}\kappa_{2,t}\kappa_{3,t}dt. \quad (\text{B.10})$$

From equations (17) and (13),

$$skew_t vol_t = 3\sigma_Y^{-2}\kappa_{3,t}. \quad (\text{B.11})$$

Therefore,

$$d[skew_t vol_t] = cons + 3\kappa_{4,t}\sigma_Y^{-4} (dY_t - \mathbb{E}_t[x_t]dt) - 9\sigma_Y^{-4}\kappa_{2,t}\kappa_{3,t}dt. \quad (\text{B.12})$$

Using equation (8),

$$dp_t - \mathbb{E}_t[dp_t] = \kappa_{2,t}\sigma_Y^{-2} (dY_t - \mathbb{E}_t[x_t]dt), \quad (\text{B.13})$$

so

$$d[skew_t vol_t] = cons + 3\frac{\kappa_{4,t}}{\kappa_{2,t}}\sigma_Y^{-2} (dp_t - \mathbb{E}_t[dp_t]) - 9\sigma_Y^{-4}\kappa_{2,t}\kappa_{3,t}dt. \quad (\text{B.14})$$

Equations (17) and (20) imply

$$3\frac{\kappa_{4,t}}{\kappa_{2,t}}\sigma_Y^{-2} = \text{sign}(skew_t)skewvol_t \left(\frac{1}{3}skew_t \right), \quad (\text{B.15})$$

while

$$9\sigma_Y^{-4}\kappa_{2,t}\kappa_{3,t} = 3\sigma_Y^{-1}skew_t vol_t^2. \quad (\text{B.16})$$

Hence,

$$d[skew_t vol_t] = cons + \text{sign}(skew_t)skewvol_t \left(\frac{1}{3}skew_t \right) (dp_t - \mathbb{E}_t[dp_t]) - 3\sigma_Y^{-1}skew_t vol_t^2 dt. \quad (\text{B.17})$$

Equation (19) implies

$$vol_t = \sigma_Y^{-2}|\kappa_{3,t}|. \quad (\text{B.18})$$

As long as $\kappa_{3,t} \neq 0$ and does not change sign locally,

$$d[vol_t] = \sigma_Y^{-2} \text{sign}(\kappa_{3,t})d\kappa_{3,t}. \quad (\text{B.19})$$

Using equation (10) and the assumption that x has stationary and independent increments,

$$d[vol_t] = cons + \sigma_Y^{-4} \text{sign}(\kappa_{3,t})\kappa_{4,t} (dY_t - \mathbb{E}_t[x_t]dt) - 3\sigma_Y^{-4}\kappa_{2,t}|\kappa_{3,t}|dt. \quad (\text{B.20})$$

Equation (8) implies

$$d[vvol_t] = cons + \text{sign}(\kappa_{3,t}) \frac{\kappa_{4,t}}{\kappa_{2,t}} \sigma_Y^{-2} (dp_t - \mathbb{E}_t[dp_t]) - 3\sigma_Y^{-4} \kappa_{2,t} |\kappa_{3,t}| dt. \quad (\text{B.21})$$

While equations (19), (20), and (13) imply

$$\text{sign}(\kappa_{3,t}) \frac{\kappa_{4,t}}{\kappa_{2,t}} \sigma_Y^{-2} = \text{sign}(skew_t) \frac{\frac{1}{3} skewvol_t vvol_t}{vol_t}, \quad (\text{B.22})$$

and

$$3\sigma_Y^{-4} \kappa_{2,t} |\kappa_{3,t}| = 3\sigma_Y^{-1} vol_t vvol_t. \quad (\text{B.23})$$

Therefore,

$$d[vvol_t] = cons + \text{sign}(skew_t) \frac{\frac{1}{3} skewvol_t vvol_t}{vol_t} (dp_t - \mathbb{E}_t[dp_t]) - 3\sigma_Y^{-1} vol_t vvol_t dt. \quad (\text{B.24})$$

C 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 $p_t = \mathbb{E}_t[x_t]$, we set $p_t \equiv \log \mathbb{E}_t[\exp(x_t)]$. We assume that the posterior cumulant generating function is analytic on a neighborhood of the closed unit disk and set $a_t = 0$ for simplicity. Then we obtain the following counterpart to corollary 1.

Proposition C.1 *vol_t follows a diffusion satisfying*

$$\begin{aligned} d(vol_t) = & \frac{1}{\sigma_Y} \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \sum_{j=1}^k \binom{k}{j} B_{k-j}(-\kappa_{1,t}, \dots, -\kappa_{k-j,t}) \mathbb{E}_t[d(x_t^j)] \\ & + \frac{1}{\sigma_Y^3} \sum_{k=2}^{\infty} \frac{\kappa_{k+1,t}}{(k-1)!} (dY_t - \mathbb{E}_t[x_t] dt) \\ & - \frac{1}{2\sigma_Y^3} \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \sum_{j=2}^k \binom{k}{j-1} \kappa_{j,t} \kappa_{k-j+2,t} dt. \end{aligned} \quad (\text{C.1})$$

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 3 is

$$\frac{\text{cov}(dp_t, dvol_t)}{\text{var}(dp_t)} = \frac{\sum_{k=2}^{\infty} \frac{\kappa_{k+1,t}}{(k-1)!}}{\sum_{k=1}^{\infty} \frac{\kappa_{k+1,t}}{k!}} \sigma_Y^{-1}, \quad (\text{C.2})$$

and the equation in proposition 4 then becomes

$$\lim_{\Delta t \downarrow 0} skew_t(p_{t+\Delta t}) (\Delta t)^{-1/2} = 3 \frac{\sum_{k=2}^{\infty} \frac{\kappa_{k+1,t}}{(k-1)!}}{\sum_{k=1}^{\infty} \frac{\kappa_{k+1,t}}{k!}} \sigma_Y^{-1}. \quad (\text{C.3})$$

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 2 holds without any changes.

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

Corollary C.1 *The analog to equation (27) is*

$$\begin{aligned} d(vol_t) = & \left(\frac{1}{3} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t}) (dp_t - \mathbb{E}_t[dp_t]) \\ & - vol_t^2 \left(\frac{\kappa_{2,t}}{\sigma_Y^2 vol_t} + \left(\frac{1}{3} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t}) \right) dt \\ & + \frac{1}{\sigma_Y} \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \sum_{j=1}^k \binom{k}{j} B_{k-j}(-\kappa_{1,t}, \dots, -\kappa_{k-j,t}) \mathbb{E}_t[d(x_t^j)]. \end{aligned} \quad (C.4)$$

The first term is identical to the first term in regression (27). In other words, the $p_t = \mathbb{E}[x_t | Y^t]$ approximation has no bearing for the relationship between the leverage effect and return skewness. The second and third terms are different from those in (27). 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

$$\begin{aligned} d(vol_t) \approx & \left(\frac{1}{3} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t}) dp_t - \frac{1}{\sigma_Y} \left(1 + \sigma_Y \left(\frac{1}{6} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t}) \right) vol_t^2 dt \\ & + \text{constant} \cdot dt, \end{aligned} \quad (C.5)$$

which is identical to (27), except for the $\sigma_Y \left(\frac{1}{6} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t})$ correction in the coefficient of vol_t^2 . The $p_t = \mathbb{E}_t[x_t]$ approximation introduces a Jensen's inequality error term that has an expansion in the cumulants; $\sigma_Y \left(\frac{1}{6} \Delta t^{-1/2} \right) skew_t(p_{t+\Delta t})$ is the leading term in that expansion.

D Accommodating the information structure in an equilibrium model

This section has two parts. First, it shows how to derive equations (4) and (5) in an endowment 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.

D.1 An endowment economy

This subsection gives an example that generates the pricing and information structure in section 3 in an endowment economy. We start with an otherwise arbitrary endowment economy with a strictly positive state-price density M_t .¹ Let D_t be an exogenous dividend rate of a redundant asset. We assume that the process for D is independent of the process for M . The assumption that D and M are independent is particularly natural for a redundant asset. The state-price density M is determined in the underlying endowment economy, while the dividend stream D can be viewed as the payoff of a financial claim that is in zero net supply and whose cash flows can be replicated using securities already traded in the economy. Introducing such a claim does not change aggregate consumption or the state-price density. This allows the uncertainty governing D to be specified independently of the uncertainty governing M .

Let W be a standard Brownian motion that is independent of both D and M . Let $\theta_t = (D^\infty, W^t)$ denote a rich information set, where D^∞ denotes the entire path of the dividend process and W^t the history of W through t . Define

$$X_t = \mathbb{E} \left[\int_{s=0}^{\infty} \frac{M_{t+s}}{M_t} D_{t+s} ds \mid \theta_t \right]. \quad (\text{D.1})$$

Since θ_t contains no information about the state-price density,

$$X_t = \int_{s=0}^{\infty} \mathbb{E} \left[\frac{M_{t+s}}{M_t} \right] \mathbb{E} [D_{t+s} \mid \theta_t] ds. \quad (\text{D.2})$$

Suppose $X_t > 0$ and let $x_t = \log X_t$. Define the signal process by

$$dY_t = x_t dt + \sigma_Y dW_t. \quad (\text{D.3})$$

Since x_s is a function of the dividend path and W^t contains W_s for every $s \leq t$, the entire history Y^t is automatically contained in θ_t .

Let $\mathcal{I}_t$ denote a filtration that only has information about the discount-factor process. Suppose the traders price the dividend stream using $\mathcal{I}_t$ together with the information in Y^t , so that

$$P_t = \mathbb{E} \left[\int_{s=0}^{\infty} \frac{M_{t+s}}{M_t} D_{t+s} ds \mid \mathcal{I}_t, Y^t \right]. \quad (\text{D.4})$$

We impose the following restriction:

$$\mathbb{E} \left[\frac{M_{t+s}}{M_t} \mid \mathcal{I}_t \right] = \mathbb{E} \left[\frac{M_{t+s}}{M_t} \right] \quad (\text{D.5})$$

¹We take the existence of M_t as given. This property holds under the usual no-arbitrage conditions. Complete markets are sufficient, but are stronger than necessary for the argument below.

for every $s \geq 0$. Independence of the discount-factor and dividends then implies

$$\begin{aligned} P_t &= \int_{s=0}^{\infty} \mathbb{E} \left[\frac{M_{t+s}}{M_t} \mid \mathcal{I}_t \right] \mathbb{E} [D_{t+s} \mid Y^t] ds \\ &= \int_{s=0}^{\infty} \mathbb{E} \left[\frac{M_{t+s}}{M_t} \right] \mathbb{E} [D_{t+s} \mid Y^t] ds. \end{aligned} \quad (\text{D.6})$$

On the other hand, since Y^t is contained in θ_t , equation (D.2) and the law of iterated expectations imply

$$\begin{aligned} \mathbb{E} [X_t \mid Y^t] &= \int_{s=0}^{\infty} \mathbb{E} \left[\frac{M_{t+s}}{M_t} \right] \mathbb{E} [\mathbb{E} [D_{t+s} \mid \theta_t] \mid Y^t] ds \\ &= \int_{s=0}^{\infty} \mathbb{E} \left[\frac{M_{t+s}}{M_t} \right] \mathbb{E} [D_{t+s} \mid Y^t] ds. \end{aligned} \quad (\text{D.7})$$

Comparing the above expression with (D.6) yields

$$P_t = \mathbb{E} [X_t \mid Y^t]. \quad (\text{D.8})$$

Since $X_t = \exp x_t$,

$$p_t \equiv \log P_t = \log \mathbb{E} [\exp x_t \mid Y^t]. \quad (\text{D.9})$$

Under the approximation in assumption 1, we set

$$p_t = \mathbb{E} [x_t \mid Y^t]. \quad (\text{D.10})$$

Equations (D.1), (D.3), and (D.10) therefore give the pricing and information structure used in section 3.

Condition (D.5) can arise in several ways. If $\mathcal{I}_t$ contains all available information about discount factors, it requires the equilibrium term structure to be deterministic, even though M itself may be stochastic. Alternatively, the traders in this market may not observe the information that forecasts discount factors, or may not use that information in valuing the asset. In these latter cases, $\mathcal{I}_t$ represents the information about discount factors that enters their valuation rather than all information available in the economy. The same information restriction must then apply to the securities that make the asset redundant, and traders using richer information must not be able to arbitrage across the two pricing environments, as under market segmentation or trading constraints.

D.2 General setup

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

$$P_t = \mathbb{E} \left[\int_0^{\infty} \frac{M_{t+s} D_{t+s} ds}{M_t} \mid \mathcal{F}_t \right], \quad (\text{D.11})$$

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

$$X_t \equiv \mathbb{E} \left[\int_0^\infty \frac{M_{t+s} D_{t+s} ds}{M_t} \mid \mathcal{G}_t \right], \quad (\text{D.12})$$

$$dY_t = x_t dt + \sigma_Y dW_t, \quad (\text{D.13})$$

where $x_t = \log X_t$, 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 $\mathcal{G}_t$,

$$P_t = \mathbb{E} [X_t \mid \mathcal{F}_t] \quad (\text{D.14})$$

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 (D.14) where $\mathcal{F}$ is the filtration induced by the Y in (D.13).

E Construction of the conditional physical moments

This appendix describes the risk adjustment procedure used throughout the empirical analysis. The principle is the same for every moment: a realized moment measured over the following month is projected on the corresponding option-implied moment (and, in the alternative specification, on additional predictors), and the fitted value of that projection is the date- t conditional physical moment.

Throughout, one “month” is $H = 21$ trading days and a year is 252 trading days. Two specifications are estimated for every moment. The *implied-moment-predictors-only* specification uses only the option-implied moment as predictor. The *additional-predictors* specification augments each projection with additional conditioning variables, described in section E.4 below.

E.1 Option-implied inputs

For the S&P 500, the option-implied inputs are, at each date t :

Implied volatility. The CBOE VIX. The daily-unit volatility state is $VIX_t/\sqrt{252}$ (where VIX is in percentage points / 100) and the implied one-month variance is

$$V_t \equiv VIX_t^2 \cdot \frac{21}{252}. \quad (\text{E.1})$$

Implied skewness. The CBOE SKEW index, converted to a standardized risk-neutral skewness via the index definition, $skew_t^Q = (100 - SKEW_t)/10$. The implied one-month *third moment* in level units is

$$ITM_t \equiv skew_t^Q \cdot \left(VIX_t \sqrt{\frac{21}{252}} \right)^3. \quad (\text{E.2})$$

Implied volatility of volatility. The CBOE VVIX. The daily-unit volatility-of-volatility state is

$$vvol_t^Q \equiv \frac{VVIX_t}{100\sqrt{252}} \cdot \frac{VIX_t}{\sqrt{252}}, \quad (\text{E.3})$$

(multiplication by the VIX is because $VVIX$ is a VIX-type index for the volatility of the log return on VIX futures). The implied one-month variance of volatility is $W_t \equiv 21 \cdot (vvol_t^Q)^2$.

Implied skewness of volatility. A Bakshi–Kapadia–Madan-style skewness of the VIX extracted from VIX options, interpolated to a constant 30-day maturity ($skewO_t$). The methods used are those from [Dew-Becker \(2024\)](#). The implied third moment of the VIX is $skewO_t \cdot W_t^{3/2}$.

For natural gas, the inputs are the constant-maturity option-implied volatility and skewness from CME natural-gas options, defined analogously to the VIX and SKEW inputs, together with the constant-maturity futures price. The natural-gas return is the day-over-day percent change of the constant-maturity futures price. No exchange-traded volatility derivatives exist for natural gas, so only the volatility and skewness are constructed for that market.

E.2 Realized moments

Realized variance. The realized one-month return variance is the sum of squared daily log returns,

$$RV_{t+1} = \sum_{j=1}^H r_{t+j}^2, \quad r_s = \log(1 + R_s). \quad (\text{E.4})$$

Realized third moment of returns. Realized skewness is measured with the model-free realized third moment of Neuberger (2012). For a fixed horizon, Neuberger shows that

$$m_{t+1}^{(3)} = \sum_{j=1}^H \left[K(r_{t+j}) + 3 \Delta v_j^E (e^{r_{t+j}} - 1) \right], \quad K(r) = 6 (re^r - 2e^r + r + 2), \quad (\text{E.5})$$

is an unbiased measure of the third moment of the cumulative return, where v^E is the *entropy* implied variance of the remaining life of the window. Because the VIX is built from the log contract, VIX^2 measures the log variance v^L , not v^E . Neuberger's equation (14) links the two through the implied third moment, $ITM = 3(v^E - v^L)$, so the entropy variance is constructed as

$$v_t^E = V_t + \frac{1}{3} ITM_t. \quad (\text{E.6})$$

In addition, equation (E.5) requires the implied variance of the *remaining life* of a fixed window $[t, t+H]$, whereas the data are constant-maturity (always H -day) variances. The fixed-horizon variance path is therefore synthesized by flat term-structure extrapolation: on day $t+j$ of the window, the variance of the remaining $H-j$ days is approximated as

$$v_j = \frac{H-j}{H} v_{t+j}^E, \quad j = 0, \dots, H, \quad (\text{E.7})$$

so that a variance shock on day j is weighted by the fraction of the window it still affects, and one day of variance rolls off deterministically each day. The increments $\Delta v_j^E = v_j - v_{j-1}$ enter equation (E.5).

Realized variance of volatility. The realized one-month variance of volatility is

$$RW_{t+1} = \sum_{j=1}^H (\Delta vol_{t+j}^P)^2, \quad (\text{E.8})$$

where vol^P is the risk-adjusted daily volatility constructed in section E.3.

Realized third moment of volatility. For the third moment of changes in volatility (an arithmetic rather than logarithmic quantity), Neuberger's arithmetic version applies:

$$m_{t+1}^{(3),vol} = \sum_{j=1}^H \left[(\Delta X_{t+j})^3 + 3 \Delta V_j \Delta X_{t+j} \right], \quad (\text{E.9})$$

where X is the raw (risk-neutral) daily volatility $VIX/\sqrt{252}$ and the variance path V_j is the synthetic fixed-horizon path (E.7) applied to the implied variance of volatility W_t .

E.3 Projections: implied-moment predictors only

Variances. The realized variances are projected on the corresponding implied variances with a Gamma generalized linear model:

$$\mathbb{E}_t [RV_{t+1}] = a + b V_t, \quad \mathbb{E}_t [RW_{t+1}] = a + b W_t. \quad (\text{E.10})$$

We use a Gamma GLM because that respects the fact that realized variance is nonnegative and highly positively skewed. In general one will expect a and b in both cases to be close to 0 and 1, respectively, and that the conditional expectations will remain positive (but see below for a setup that avoids that potential issue).

The conditional physical volatilities in daily units are rescaled fitted values (where $\widehat{RV}_t$ denotes the forecast based on time- t information):

$$vol_t^P = \sqrt{\widehat{RV}_t/21}, \quad vvol_t^P = \sqrt{\widehat{RW}_t/21}. \quad (\text{E.11})$$

Third moments. The realized third moments are projected on the implied third moments,

$$m_{t+1}^{(3)} = a + b \cdot ITM_t + \varepsilon_t, \quad m_{t+1}^{(3),vol} = a + b \cdot skewOW_t^{3/2} + \varepsilon_t, \quad (\text{E.12})$$

estimated by maximum likelihood with Student- t errors with five degrees of freedom rather than least squares: realized third moments are extremely fat-tailed, and a handful of crisis windows would otherwise dominate the fit, but at the same time the realized third moment is not constrained to being only positive or negative. The likelihood is maximized on standardized variables (each variable demeaned and scaled by its standard deviation) for numerical conditioning, and the fitted values are mapped back to the original units.

The fitted third moments are then converted to standardized skewnesses using the estimated physical conditional variances from the previous step,

$$skew_t^P = \frac{\widehat{m}_t^{(3)}}{(\widehat{RV}_t)^{3/2}}, \quad skewOW_t^P = \frac{\widehat{m}_t^{(3),vol}}{(\widehat{RW}_t)^{3/2}}. \quad (\text{E.13})$$

Both the numerator (a fitted physical third moment) and the denominator (a fitted physical variance) are objects under the physical measure, so the ratio is a physical standardized skewness.

E.4 Projections: additional predictors

The second specification, which includes multiple additional predictors, changes the projections in two ways. First, the *variance* projections switch to a Gamma GLM with log link on

the logged implied moment,

$$\mathbb{E}[RV_{t+1}] = \exp(a + b \ln V_t + c \ln RV_t^{trail} + d CS_t + e CAPE_{t-1mo}), \quad (\text{E.14})$$

a power-law specification whose fitted values are positive by construction, to avoid negative fitted values. Second, every projection (variances and third moments) adds the same three predictors:

- the projection's own trailing realized moment: for the variance fits, the trailing 63-day realized variance rescaled to a 21-day equivalent (RV_t^{trail}); for the third-moment fits, the average of the three most recent *completed* 21-day windows of the same realized third-moment measure (entered in levels, not logs, since third moments are signed). Whole 21-day windows are used because the Neuberger measure is defined window-by-window — its drift term uses fixed-horizon weights — and is therefore not additive across days the way realized variance is;
- the Baa–Aaa corporate credit spread on day t (CS_t);
- the cyclically adjusted price–earnings ratio (CAPE), lagged one month.

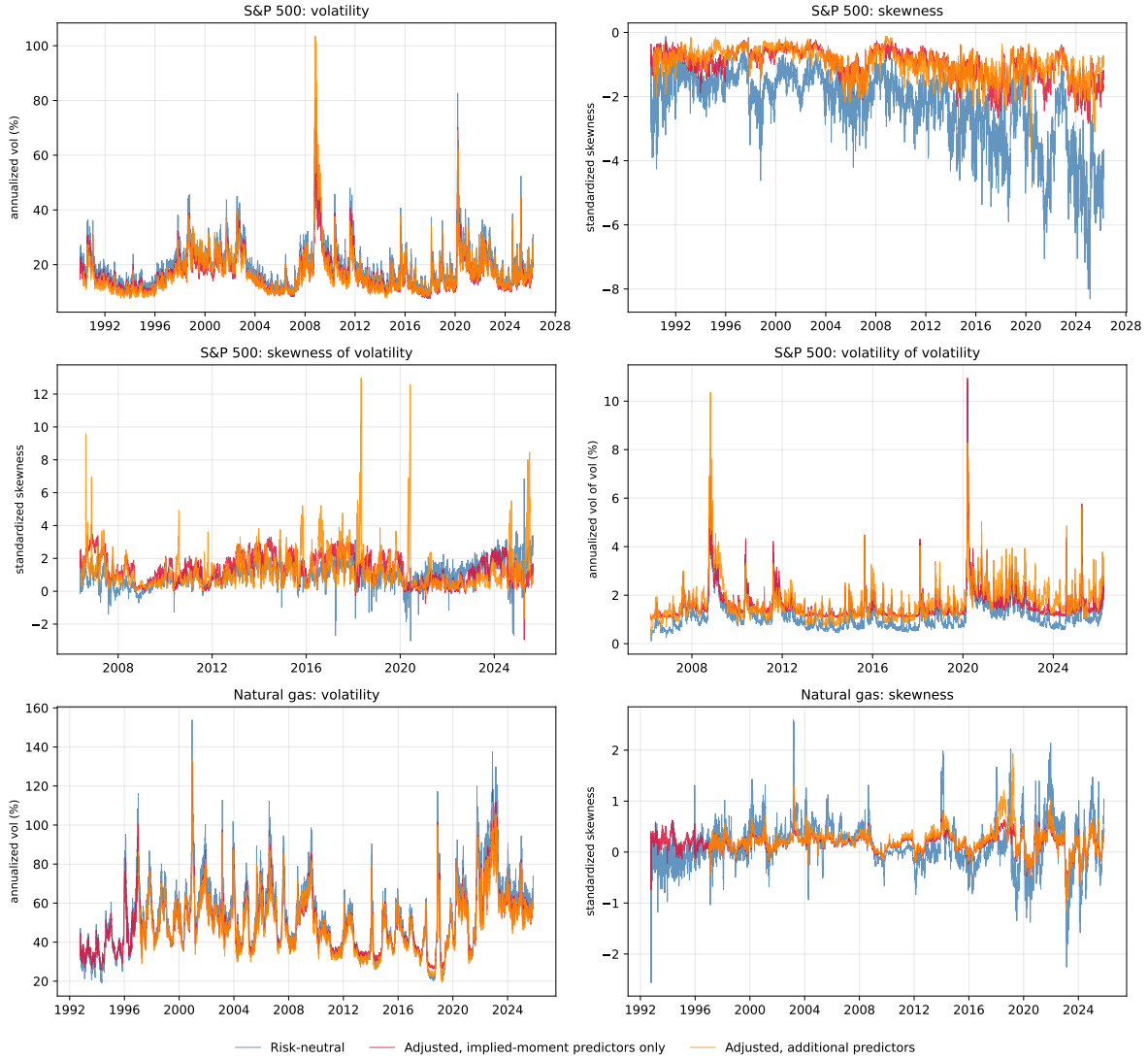
The third-moment projections keep the standardized Student- $t(5)$ maximum likelihood estimator; only their predictor set changes.

For natural gas the same two projections (volatility and skewness) add one further predictor: the log futures–spot basis, $\ln(F_t/S_t^{HH})$, where F is the constant-maturity futures price and S^{HH} the FRED Henry Hub spot price. The spot series begins in 1997, so the natural-gas additional-predictors series start then.

E.5 The risk-adjusted series

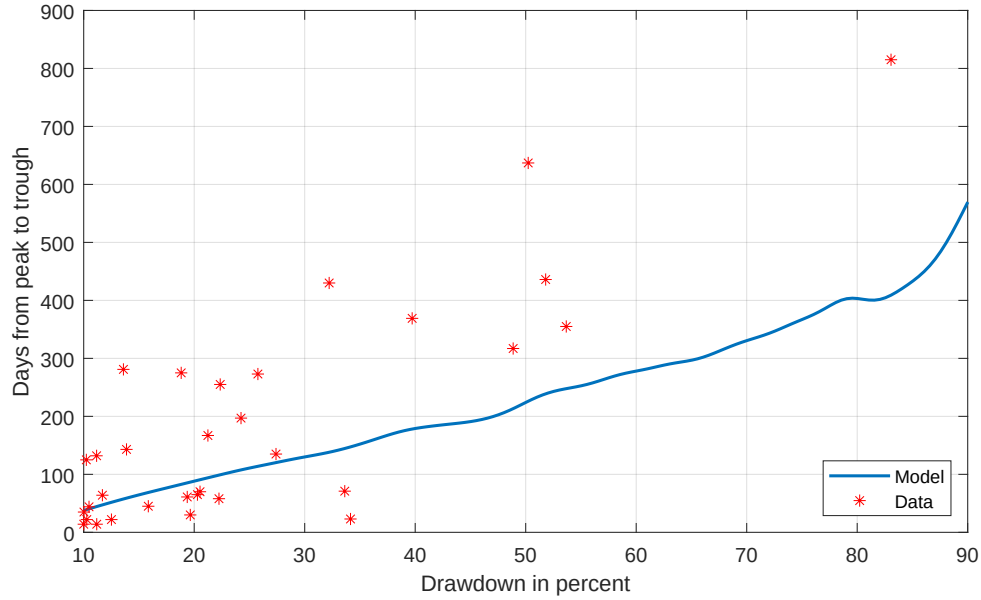
Figure E.1 plots each risk-neutral series against its two risk-adjusted counterparts: the four S&P 500 moments (volatility, skewness, skewness of volatility, volatility of volatility) and the two natural-gas moments (volatility, skewness).

Figure E.1: Risk-neutral and risk-adjusted conditional moments



Note: Each panel overlays the option-implied (risk-neutral) conditional moment with its two risk-adjusted counterparts: the projection on the implied moment only, and the projection with additional predictors (own trailing realized moment, Baa–Aaa credit spread, one-month-lagged CAPE and, for natural gas, the log futures–spot basis, which starts that specification in 1997). Volatility and volatility of volatility are plotted as annualized percentages (daily level $\times \sqrt{252} \times 100$).

Figure E.2: Time from peak to trough in large drawdowns



Note: The blue line plots the average number of days from peak to trough in the model simulations for a drawdown (decline relative to the running maximum) of a given size. The average is calculated using a kernel smoother on the simulated realized drawdowns. The red stars represent the time from peak to trough for drawdowns observed in the US stock market.

Because the behavior of drawdowns depends not only on the equity premium but also the risk-free rate, we use the following procedure to put the data and simulation on the same footing. For the data, drawdowns are calculated based on cumulative adjusted returns, where the adjusted return is defined as the excess return plus 3.3% per year, which is the average risk-free rate in the sample. In the model, we set the mean growth rate of prices to be equal to 10.9%, which corresponds to the mean of the equity premium plus the mean risk-free rate (i.e. the mean total return on equities, corresponding to g in the model). That ensures that both series have the same average growth rate which makes their drawdowns comparable.