The dependence of belief dynamics on beliefs: implications for stock returns

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

October 13, 2024

Abstract

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

1 Introduction

Motivation

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

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

has over a century of data on expectations, in modern times at frequencies that are nearly continuous, and furthermore data on options gives measures of higher-order moments of beliefs.[1] This paper's basic goal is to understand the dynamics of beliefs in a general model of information acquisition, and in particular to understand the relationship between the various conditional moments of beliefs.

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

Beyond that first relationship, there are numerous other features of returns to understand – negative skewness and positive excess kurtosis in returns, positive skewness and excess kurtosis along with long memory in the volatility of returns, and a strong relationship between conditional skewness and the future covariance between prices and volatility. No other dataset provides the same granularity when testing a model of beliefs.

Contribution

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

The theoretical structure is built around the idea that agents fundamentally want to know the discounted value of a security's cash flows. There are many ways that cash flows and information can be modeled, but they are all broadly asking how expectations are updated as information arrives. We therefore study a simple but general setup: the NPV follows some arbitrary process, and agents continuously receive signals about it, which represent in reduced form the aggregate of all the information people observe in reality. Since the NPV

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

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

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

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

process is essentially unconstrained, the analysis nests a wide range of specifications that have been studied in the literature.

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

The theorem is useful because it shows that the conditional moments are in fact sufficient statistics, measuring how beliefs respond to signals. The paper's core motivating fact is that aggregate stock returns are negatively correlated with innovations to volatility. When agents are learning about fundamentals, that happens *if and only if* agents' conditional third moment for fundamentals is negative, and in fact the magnitude of the return-volatility relationship is proportional to the conditional third moment, representing a strong and testable prediction.

Beyond sensitivity, the theorem also yields expressions for the drift in the cumulants. Interestingly, mean-reversion in volatility is generically nonlinear: instead of depending on the current level of volatility, the drift in volatility is proportional to its *square*. That fact yields the sort of long memory that has been observed in stock market volatility. Rather than requiring a highly complicated model, or even any particular assumptions about fundamentals at all, long memory is an inevitable feature of information acquisition.[5]

Long memory also represents a prediction that is distinct from what would be implied by a model in which the leverage effect is driven purely by risk premia. A second prediction that also distinguishes the model is that the magnitude of the leverage effect should be related to the current conditional skewness of returns, and with a specific value for the coefficient, which we confirm in the data.

After developing a few more theoretical results, including for fragility of beliefs with respect to errors in agents' signals, and examining some specific examples, we move on to a quantitative analysis of the model's predictions. First, we examine a very simple paramet-

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

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

Second, and perhaps more relevantly, we show how to derive *nonparametric* predictions from the model – tests and estimates that can be obtained without knowing anything about the underlying dynamic process for fundamentals. First, the model has predictions for the relationship between volatility its own lag, returns, and skewness, that we test and find hold well in the data. Second, it is possible to estimate agents' implied uncertainty about the level of fundamentals without knowing the underlying model. In US stock market data, we estimate that uncertainty to have a standard deviation of between 11 and 18 percent. In a survey administrated by Yale University since the 1980's, cross-sectional disagreement about the fundamental value of the stock market has a standard deviation of 17 percent, which provides some independent support for our estimate (subject to the usual caveat that disagreement and uncertainty are theoretically distinct).

Stepping back

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

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

Past work

As discussed above, this paper is most closely related to past work studying non-Gaussian

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

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

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

Outline

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

2 Model setup and solution

We motive the analysis in terms of asset prices, and this section begins by describing the simple asset pricing framework we study. That setup is designed, to lead to a standard filtering problem. The solution to that problem – in theorem 1 – is general, and not actually dependent on the asset pricing setting.

2.1 Model setup

[7] The class of Feller processes with finite-dimensional states is extremely broad and incorporates all models commonly studied in economics – jump diffusions, Markov chains (e.g. regime-switching processes), semimartingales, affine processes, etc. θ_t need not be stationary at any order, except to the extent required for asset prices to exist.

2.1.1 Dynamics of fundamentals

There is a latent variable θ_t representing the state of the economy. If agents knew θ_t, we assume they would set log stock prices equal to

$$x_t \equiv \log E\left[\int_{s=0}^{\infty} \frac{C_{t+s}}{R_{t,t+s}} ds \mid \theta_t\right] \tag{1}$$

where C_t denotes the time-t cash-flow and $R_{t,t+s}$ is the discount rate between dates t and $t+s$. x_t is the object of interest – we think of it as the (log) fundamental value that stocks would take on if agents had full information in the sense of being able to measure the true state of the economy, θ_t. We make no particular assumptions about risk aversion. x_t is simply what agents would pay if they knew the true state of the economy.

In terms of restricting, instead of making assumptions about θ_t and how it determines cash-flows and discount rates, we work with x_t directly. A structural model will have implications for the dynamics of C_t and $R_{t,t+s}$, and one can then ask whether the x_t process they induce satisfy our assumptions. The necessary restrictions on x_t are reported in appendix A.1.1. The key requirement is that x_t, in reduced form, is driven by a Markov process with the Feller property. Examples of such processes are Brownian motions, jump diffusions, and Levy-stable processes – in short, anything studied in economics. There is no requirement that x_t is stationary at any order, and all of its moments need not exist (i.e. it may be heavy-tailed). x_t additionally can be the outcome of a learning model. For example, cash-flows might have a mean growth rate that is unknown, and then θ_t includes the entire history of cash-flows as part of the current state. Essentially what the Feller assumption rules out is just that the dynamics of x_t are discontinuous in terms of the state. For example, the mean of x_t cannot shift discretely when the state θ_t crosses some boundary.

[7] An example of a violation of the Feller property is if the dynamics of the process change discontinuously with respect to the state.

In principle, the results hold when θ_t is an element of a much more general class of topological spaces. For example, under certain assumptions, the results here also hold when the state θ_t is itself a function (e.g. a cross-sectional distribution).

For the purposes of the present paper, though, θ_t being a finite-dimensional vector is a sufficient level of generality.

2.1.2 Information flows

We denote the log of stock prices by p_t. If agents observed the true state, θ_t, they would set prices as $p_t = x_t$. In the absence of knowing θ_t we simply assume that agents directly receive a signal about the NPV – the actual value relevant for prices – which follows

$$dY_t = x_t dt + \sigma_{Y,t} dW_t \tag{2}$$

where $\sigma_{Y,t}$ follows some exogenous process.[8] A reasonable benchmark is that $\sigma_{Y,t}$ is constant, but it could also vary over time, giving a form of time-varying uncertainty (we provide evidence below for the US stock market that it appears stable). Obviously this is not the only possible information structure. Agents could receive signals about nonlinear functions of x_t, such as its moments, or about θ_t, which might contain relevant information about the future path of x_t. However, given that x_t is what agents ideally would like to know, it makes a certain amount of sense to assume that it what agents learn about. Furthermore, this information specification nests full information as $\sigma_{Y,t} \to 0$.

We then assume that

$$p_t = E\left[x_t \mid Y^t\right] \tag{3}$$

where Y^t is the history of signals up to date t.[9]

The assumption that information flows diffusively matters for the analysis, but it is not completely restrictive – section 4.3 discusses how the results apply when information may arrive discretely or when there are large information revelation events. The assumption that Y_t is informative about x_t itself and not either other features of θ_t or other functions of x_t (e.g. it could depend on x_t^2) is also a restriction. For an agent who is trying to price assets, x_t is a sufficient statistic – it is the ideal price they would like to pay. However, in reality agents certainly might receive information about other features of the world. Our analysis is extremely general in the dynamics for fundamentals, but pays for that generality with a tight restriction on the information structure.

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

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

[9]Thiere is perhaps a slight conceptual inconsistency here, in that we're saying $p_t = \log E[\log E[NPV_t \mid \theta_t] \mid Y^t]$. The appearance of the logs here is entirely to make the model behave naturally when x_t follows a geometric process. The Campbell–Shiller approximation is an alternative way to get to the same place. If, instead, x_t follows a linear process, then the logs are unnecessary and we simply have nested expectations.

of signals. None of this needs to be about asset prices.

2.2 Solution to the filtering problem

We describe the dynamics of beliefs about fundamentals by describing the dynamics of the agents' posterior cumulants. Formally, the cumulants are the derivatives of the log characteristic function, but for the purposes of this paper we will only focus on the first three, which are equal (for all distributions for which they exist) to the first three central moments. In other words, almost nothing will be lost if the word "moment" is substituted for "cumulant" in everything that follows.

Denote the n-th cumulant of the time-t conditional distribution of x_t by $\kappa_{n,t}$.[10] Since the first cumulant is the expectation, $p_t = \kappa_{1,t}$.

Theorem 1 *Given (2) and restrictions on x_t given in appendix A.1.1,, for all n for which the $n+1th$ cumulant exists*[11]

$$d\kappa_{n,t} = \frac{\kappa_{n+1,t}}{\sigma_{Y,t}^2} \left(dY_t - E_t\left[x_t\right] dt\right) - \frac{1}{\sigma_{Y,t}^2} \sum_{j=2}^{n} \alpha_j^{(n)} \kappa_{j,t} \kappa_{n-j+2,t} dt + E_t\left[d\left(x_t^k\right)\right] \tag{4}$$

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

That result follows from a straightforward application of results in Lipster and Shiryaev (2013) and Bain and Crisan (2009).[12] The most valuable feature of theorem 1 is that it shows that the sensitivity of each cumulant to signals is proportional to the current value of the next cumulant.

The first three cumulants, since they map into the first three moments, are worth writing out directly.

[10] The first three cumulants are equal to the first three central moments. For the only the normal distribution all higher-order cumulants are equal to zero.

[11] Since the cumulants are derivatives of a function, if $\kappa_{n+1,t}$ exists then all lower-order cumulants also exist. Note that the posterior of x_t conditional on y^t is necessarily subgaussian, meaning that all moments and cumulants exist (Guo et al. (2011)). However, the step from there to a stochastic differential equation for the posterior cumulants may be nontrivial.

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

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

$$dp_t = d\kappa_{1,t} = \frac{\kappa_{2,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[dx_t\right] \quad (5)$$

$$d\,\mathrm{var}_t\left[x_t\right] = d\kappa_{2,t} = \frac{\kappa_{3,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{\kappa_{2,t}^2}{\sigma_{Y,t}^2}dt \quad (6)$$

$$dE_t\left[\left(x_t - E_t\left[x_t\right]\right)^3\right] = d\kappa_{3,t} = \frac{\kappa_{4,t}}{\sigma_{Y,t}^2}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^3\right)\right] - \frac{3}{\sigma_{Y,t}^2}\kappa_{2,t}\kappa_{3,t}dt \quad (7)$$

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

The intuition for the result is surprisingly simple: in (4), the gain, $\kappa_{n+1,t}/\sigma_{Y,t}^2$, is a local regression coefficient. For the mean in equation (5), for example, the numerator $\kappa_{2,t}$ is equal to $\mathrm{cov}_t\left(x_t, dY_t\right)/dt$ and the denominator is equal to $\mathrm{var}_t\left(dY_t\right)/dt$, so their ratio is exactly the regression coefficient. Similarly, $\kappa_{3,t} = E\left[\left(x_t - E_t\left[x_t\right]\right)^3\right]$ is equal to $\mathrm{cov}_t\left(\left(x_t - E_t\left[x_t\right]\right)^2, dY_t\right)/dt$, so $\kappa_{3,t}/\sigma_{Y,t}^2$ is again a regression coefficient.

3 Predictions

We now examine the predictions of theorem 1 for the behavior of returns, beginning with volatility. Recall that the **leverage effect** is the name for the observed negative correlation between innovations to return volatility and returns themselves. The key theme underlying the predictions is that richness in the joint dynamics of higher moments can be a simple consequence of learning, without requiring a complicated model.

3.1 Volatility and the leverage effect

For stocks, at high frequency cash flows are predetermined, and in any case the variance of changes in cash flows for the aggregate US stock market at even the monthly frequency is tiny compared to changes in prices.[13] We therefore treat return volatility as equal to price volatility. Formalizing the discussion above, we have

[13] The historical variance of monthly returns 2.85×10^{-3}, while the variance of dividend growth is over 600 times smaller – 4.46×10^{-6}.

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

$$std\left(dp_t\right) = \frac{\kappa_{2,t}}{\sigma_{Y,t}} dt^{1/2} \tag{8}$$

Again, the conditional volatility of prices depends on agents' current posterior variance over fundamentals, $\kappa_{2,t}$. So, up to $\sigma_{Y,t}$, price volatility measures uncertainty. If we can measure $\sigma_{Y,t}$ (which we show below we can), then we can also measure investors' uncertainty about fundamentals, $\kappa_{2,t}$.

Proposition 1 *The instantaneous leverage effect, measured as the coefficient in a regression of changes in the conditional variance of returns on price changes is*

$$\frac{\text{cov}\left(dp_t, d\left[std\left(dp_t\right) dt^{-1/2}\right]\right)}{\text{var}\left(dp_t\right)} = \frac{\kappa_{3,t}}{\sigma_{Y,t}\kappa_{2,t}} \tag{9}$$

The leverage effect is completely determined by the second and third moments of agents' conditional distribution and the noise in agents' signals (again, a set of sufficient statistics). Since $\sigma_{Y,t}^2$ and $\kappa_{2,t}$ are strictly positive, a necessary and sufficient condition for the existence of a leverage effect is that $\kappa_{3,t} < 0$: there is a leverage effect if and only if agents' posterior distribution for fundamentals is negatively skewed. And the fact that we observe a leverage effect in nearly all months in the data, including during severe downturns, then implies that the conditional skewness is negative in essentially all states of the world (at least among those in our sample).

The intuition here is relatively simple: a negative third moment means that the right tail of the posterior is longer than the left. When agents receive good news about fundamentals, that tells them they are likely on the narrower side of the distribution, and their conditional uncertainty falls. That intuition is generic – it is not dependent on some specific specification. And, additionally, the value of theorem 1 is that it formalizes that intuition and shows that the third moment is in fact that correct measure of asymmetry in the distribution to capture such an effect.

3.2 Slow decay in volatility

The second term in (6) shows how volatility decays. When $\kappa_{2,t}$ (and hence also price volatility) is high, $\kappa_{2,t}^2 \sigma_{Y,t}^{-2} dt$ also grows, pulling volatility back down towards its steady state. Interestingly, though, unlike standard models (e.g. an AR(1) or Ornstein-Uhlenbeck process), the mean reversion is *quadratic*. That is, the rate of mean reversion increases as $\kappa_{2,t}$ moves further above its mean. When volatility moves back down, mean reversion slows.

There is a large literature studying nonlinearity in volatility dynamics in securities markets. The form of mean reversion here is consistent with that literature, in that the decay is non-exponential.[14] When jumps up in $\kappa_{2,t}$ are large relative to its steady-state value, its decay is approximately of the form $1/(a+t)$ for a coefficient a (that depends on the other parameters of the model). This is exactly the polynomial decay studied in the literature on long memory in volatility.

In other words, when investors are learning about fundamentals dynamically, long memory is to be expected. It is a simple consequence of the dynamics of second moments in filtering models. We examine the model's ability to fit detailed data on volatility dynamics in more detail in section 6.

3.3 Skewness in returns

Since P_t follows a diffusion, its instantaneous skewness is formally zero. Skewness arises as returns interact with changes in volatility. Informally, to first order in $\sigma_{Y,t}$ and at some (small) horizon Δt,

$$skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3\Delta t^{1/2}\sigma_Y}\approx\frac{\kappa_{3,t}}{\kappa_{2,t}} \tag{10}$$

That is, the conditional "instantaneous" skewness of returns again depends on the second and third moments of the posterior. As $\Delta t\to 0$, skewness goes to zero. But, locally, it scales with $skew_t\left(x_t\right)\kappa_{2,t}\sigma_{Y,t}$. That fact provides a link between indexes of the conditional skewness of returns ($skew_{t\to t+\Delta t}\left(dp_t\right)$), such as the CBOE's option-implied skewness, and the conditional skewness of fundamentals, $skew_t\left(x_t\right)$, which determines the leverage effect.

3.4 Skewness in volatility

In the data, the VIX is itself skewed. Table 1 in the quantitative analysis below shows that it true of both its level and monthly changes. The source of that effect is visible if we combine equations (6) and (10) (taking the latter as an equality here) to obtain

$$std\left(d\kappa_{2,t}\right)=\frac{\left|skew_t\left(dp_t\right)\right|}{3}\frac{\kappa_{2,t}}{\sigma_{Y,t}^2} \tag{11}$$

All else equal, the volatility of innovations to $\kappa_{2,t}$ scales with $\kappa_{2,t}$ itself. When $\kappa_{2,t}$ falls towards zero, the volatility of its innovations quickly becomes much smaller, while they grow

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

when $\kappa_{2,t}$ rises. That effect creates a long right tail in the level of $\kappa_{2,t}$ itself, and any skewness in $\kappa_{2,t}$ itself is also inherited by $d\kappa_{2,t}$.

$\kappa_{3,t}$ also plays a role in volatility of volatility through the term $skew_t(dp_t)$. Again, simple filtering gives a generically rich relationship between higher moments.

Past work (e.g. Bollerslev, Tauchen, and Zhou (2009)) has emphasized the importance of time-varying vol-of-vol. This present model gets it through an endogenous mechanism. Note also that this variation does not just come from the volatility of fundamentals following a nonlinear process, as in Cox, Ingersoll, and Ross (1985).

3.5 Summary

To briefly summarize so far, simple filtering predicts a leverage effect when $\kappa_{3,t} < 0$, long memory in volatility, skewness in returns for $\kappa_{3,t} \neq 0$, skewness in both levels and changes in volatility, and time-varying volatility of volatility. None of this requires any mechanism more complicated than Bayesian updating in the presence of nonzero higher moments.

3.6 Examples

This section briefly considers a few examples which allow us to further describe some of the model's implications.

3.6.1 Linear Gaussian process

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

As discussed above, this can involve learning. That is, the analysis does not assume that agents know all the parameters of the economy. For example, suppose the log cash-flow in period t is $dc_t = \bar{c}dt + \sigma_c dZ_t$, where Z_t is a standard Wiener process. If agents to not know $\bar{c}$, then the state variable is $\theta_t = c_t$, and $x_t = t^{-1}c_t$+constant. Everything remains linear and Gaussian. What matters is not that agents know all the economy's parameters – they just need to know its structure.

3.6.2 Markov switching process

Veronesi (1999) studies a two-state switching model in which the latent state x_t switches between a low and a high value at rates λ_{HL} and λ_{LH}, respectively, and agents have a Gaussian signal as specified above. In this case, the low and high values of x_t can be normalized to 1 without loss of generality. Such a two-state setup is also common in analyses of the business cycle, e.g. Hamilton (1989). Agents' posterior at any given time has only a single parameter, π_t, their posterior probability that $x_t = 1$.

The conditional variance and third moment of x_t, which drive price dynamics, are simple functions of π_t:

$$\kappa_{2,t} = \pi_t (1 - \pi_t) \tag{12}$$

$$\kappa_{3,t} = (1 - 2\pi_t) \times \kappa_{2,t} \tag{13}$$

The variance here then is a bell-shaped function of π_t, peaking at $1/4$ for $\pi_t = 1/2$, and declining to zero on both sides. The third moment is equal to the variance times $(1 - 2\pi_t)$. It is equal to zero for $\pi_t \in \{0, 1/2, 1\}$. For $\pi_t \in (0, 1/2)$, $\kappa_{3,t} > 0$, and for $\pi_t \in (1/2, 1)$, $\kappa_{3,t} < 0$.

Economically, when π_t is near 1 so that agents are confident they are in the good state, volatility is low, but the third moment is strongly negative, so there is a leverage effect. However, when a bad state is realized and investors have seen enough signals to be confident in that, so that π_t is near zero, the leverage effect *reverses*: agents no longer worry about the economy getting worse, so there is only upside and $\kappa_{3,t} > 0$.

These results illustrate the importance of agents continuing to learn in bad states. If learning stops once agents know the economy is in a recession, then the leverage effect disappears or even reverses. But we do not observe that empirically. One response might be to simply add more states – David and Veronesi (2013) have six states – but it will still always be the case that once agents are confident they are in the worst state, skewness must turn positive.

3.6.3 Exponentially distributed fundamentals

As discussed above, because the leverage effect holds at all times, a model that fits the data needs to have the feature that the conditional distribution of fundamentals is always skewed left. The model in this section generates much of that negative skewness and is also related to the quantitative model that we study in the next section, but stylized so that it can be solved in closed form. In particular, we assume that fundamentals, x, are drawn on date 0 from the distribution

$$x \sim \begin{cases} 0 \text{ with prob. } (1-\pi_0) \\ Exponential\,(\lambda) \text{ with prob. } \pi_0 \end{cases} \tag{14}$$

and subsequently remain fixed forever, while agents observe the continuous Gaussian signal

$$Y_t = xdt + \sigma_Y dW_t \tag{15}$$

This example is therefore useful for understanding the dynamics of learning about the state following the possibility of one extreme event. The true dynamic model can be thought of intuitively as this miniature model being run repeatedly – agents repeatedly (or continuously) learn about the state of the economy, with some belief that there may be a crash. That is, one could imagine (very approximately) that the model in this section plays out once a year, so that π_0 represents the fraction of years with a crash.[15]

This model is solvable in closed form:

Solution 4 *Having observed the signal process Y up to date t, the agent's posterior is*

$$x \sim \begin{cases} 0 \text{ with prob. } (1-\pi_t) \\ TN\left(0, \mu_t, \sigma_t^2\right) \text{ with prob. } \pi_t \end{cases} \tag{16}$$

where $TN\left(0, \mu_t, \sigma_t^2\right)$ is a normal distribution truncated above at zero and π_t, μ_t, and σ_t^2 are functions of Y_t and t.

Appendix A.1.4 gives the formulas for π_t, μ_t, and σ_t^2.

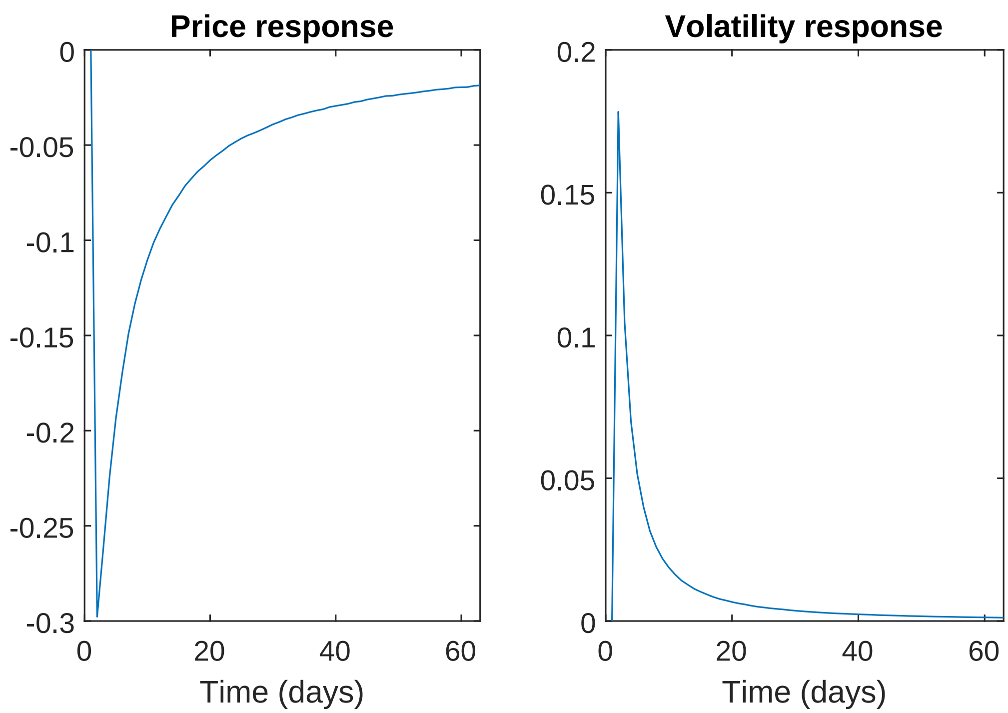
It is not the case in general that the distribution of x conditional on the signals must be negatively skewed. However, note that a normal distribution truncated above at zero is always negatively skewed. That force means that even when π_t is large x remains negatively skewed as long as σ_t^2 is not too small.

Impulse response functions This section examines two impulse response functions – to errors in the signal, and to fundamentals – in the exponential example. First, consider a negative shock to the error in dY_t, i.e. to $\sigma_{Y,t} dW_t$. The IRFs are averaged across draws from the prior distribution for the state. The parameters are chosen to match those used in the dynamic model in the next section. The figure below plots the response of prices and volatility. The negative shock to the signal lowers prices temporarily and raises volatility,

[15] This should not be taken literally. The economy does not reset on January, 1st each year. t does not represent the day of the year. The model in this section just illustrates some features of learning in a specific parametric model that happens to have a closed-form solution.

consistent with the leverage effect and the fact that beliefs are typically negatively skewed in this model. Prices then recover fairly quickly over the course of a few weeks, and volatility follows a mirror image, falling at about the same rate.

Figure 1: Response to negative error in the signal

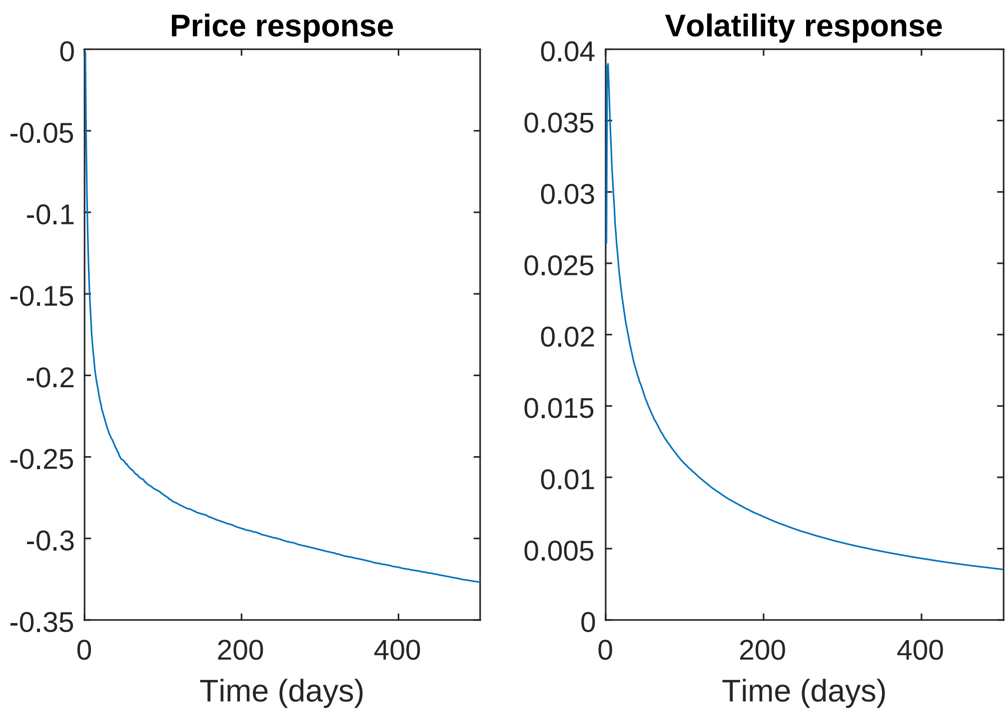


Note: The left-hand panel plots the IRF for prices – the conditional expectation of x – and the right hand for the conditional standard deviation of prices. The shock is a one-time unit standard deviation negative error in the signal (i.e. a negative realization of $\sigma_Y dW$.

The second IRF is slightly less standard. The "shock" now is the choice of x. From the agent's perspective this is a shock like any other, it's just a shock that only occurs on a single date (date 0). Informally, this can be thought of as an impulse response for the realization of a disaster. We study that more formally in the quantitative evaluation below. The IRF, then, is equal to the average path of p_t, along with its conditional volatility, for $x = -\lambda$ relative to $x = 0$.

When x has a negative realization, prices initially are no different from a positive realization, since on date 0 investors cannot yet know that anything is different. As they accumulate signals, though, they eventually come to the conclusion that fundamentals are weak, and p_t eventually converges to $-\lambda$ (which was chosen for this particular calibration – in reality every crash realization for x would take on a different value). Volatility initially rises with the accumulation of negative signals, and then starts to fall due to the fact that when uncertainty is higher, information effectively flows more quickly. In this case, note

Figure 2: Response to a negative realization of fundamentals



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

that volatility takes far longer to revert than in the case of a purely transitory shock – two years into the simulation, about 1/10th of the initial rise in volatility still remains. That demonstrates the model's long memory in volatility.

3.6.4 Are there simple and tractable special cases?

In looking at theorem 1, a natural question is whether it is possible to shut down the cumulants after some order. Unfortunately that is not strictly possible because there is no distribution such that $\kappa_n = 0$ for all n greater than some $\bar{n}$ other than the normal distribution. The normal is the special case where we can focus on just two cumulants. There is no other distribution for which it is possible to just track a finite number of moments (though that does not say how large the numerical error would be from doing so).

As an alternative, one might hope to track just a subset of the cumulants. For example, suppose agents begin with a symmetrical prior for x_t – i.e. one for which all odd cumulants are equal to zero. Unfortunately, simple inspection of theorem 1 shows again that that is not possible – the odd cumulants remain zero with probability zero, since their responses to the signal dY_t depend on the even cumulants, which are nonzero (except for the normal distribution).

4 Extensions and robustness

This section considers a few extensions of the analysis. We first discuss fragility with respect to errors in agents' signals. Second, how to use local information about prices to infer global features of investors' beliefs. Last we examine robustness of the results to discreteness in the signals and also provide a result on the limiting behavior when the noise in the signal becomes small.

4.1 Fragility

To define fragility, first we define a type of sensitivity.

Definition 5 ***Noise sensitivity*** *is the marginal response of prices to errors in the signal,* $dP_t/d\left(\sigma_{Y,t}dW_t\right)$

Since agents cannot directly distinguish errors in the signal $(\sigma_Y dW_t)$ from the part driven by fundamentals (Xdt), any sensitivity of prices to signals necessarily also leads to sensitivity to errors.

If sensitivity rises following negative shocks, then consecutive negative shocks have progressively larger effects on prices. We therefore refer to the *derivative* of noise sensitivity as fragility:

Definition 6 ***Fragility*** *is* ***minus*** *the proportional response of noise sensitivity to errors in the signal,* $\frac{-d^2P_t/d\left(\sigma_{Y,t}dW_t\right)^2}{dP_t/d\left(\sigma_{Y,t}dW_t\right)}$

Fragility can equivalently just be seen as concavity in the mapping from noise to prices. That concavity means that noise has larger downward than upward effects (to second order).[16]

Corollary 7 *Noise sensitivity and fragility are equal to:*

$$\begin{aligned} \textit{Noise sensitivity} &= \frac{\kappa_{2,t}}{\sigma_{Y,t}^2} && (17) \\ \textit{Fragility} &= -skew\left(x_t \mid Y^t\right)\kappa_{2,t}^{1/2} && (18) \end{aligned}$$

[16]Note that defining fragility based on concavity is similar to the approach advocated by Taleb and Douady (2013), but more as a local concept (whereas that paper is formally focused on tails – i.e. $-\infty$ limits). Axenie et al. (2024) define fragility similarly to here based on convexity in the response to shocks.

Fragility is (minus) the sensitivity of the gain to innovations in the signal, which theorem 1 shows depends on the conditional third moment of x_t. When the conditional skewness is more negative or the conditional volatility greater, prices are more fragile, in the sense that noise in the signal can have larger negative than positive effects. Periods of high uncertainty are thus periods where investors are learning a lot – the gain is high – but they are also more susceptible to errors than average, and (holding skewness fixed) their beliefs are most fragile.

As with all the results, this follows from Bayes' theorem. When the left tail of the posterior is relatively long ($\kappa_{3,t} < 0$) agents think there is a larger chance that fundamentals are far below than far above their current mean. That then makes them update beliefs more strongly in response to negative than to positive signals, leading to fragility.

Corollary 7 motivates the use of option-implied skewness as a real-time measure of fragility. It does not just reflect the subjective distribution of returns. It also measures the extent to which noise in investors' information can lead to large price declines. Specifically, combining (18) and (10) we have

$$\textit{Fragility} \approx -\frac{dt^{-1/2}}{3\sigma_{Y,t}} \textit{skew}_t\left(dp_t\right) \tag{19}$$

So a skewness index that measures $\textit{skew}_t\left(dp_t\right)$ is not only a measure of the conditional distribution of returns, it is also informative for how prices respond to errors.

Proposition 2 *In the exponential example (section 3.6.3), for all t and for all realizations of Y_t: fragility is positive and decreasing in π_t.*

The fact that fragility is decreasing in π_t means that it is largest when there is a very small probability of a large negative shock.

4.2 Inferring global properties from local information

Theorem 1 says that the local properties of prices are determined by sufficient statistics describing the global properties of beliefs about fundamentals. So far we have analyzed what is required of those global properties in order to generate local price behavior observed empirically. This section reverses the analysis: how can we use local information about prices to learn about global properties of beliefs?

Under theorem 1, prices follow a diffusion with volatility $\kappa_{2,t}/\sigma_{Y,t}$. There are standard results that then allow for consistent estimation of the diffusive volatility based on high-frequency observations of prices. Those methods therefore allow for real-time estimation of $\kappa_{2,t}/\sigma_{Y,t}$ without knowing any of the underlying parameters of the model. Again, $\kappa_{2,t}$ measures a global feature of beliefs – the conditional variance. So, subject to some potential contamination from $\sigma_{Y,t}$, it is possible to estimate that global feature in real time using only local variation in prices.

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

Note that the fragility studied in the previous section also has the same local/global features as volatility and vol-of-vol. We defined fragility as a local concept, but it depends on global characteristics of beliefs.

4.3 Discrete information revelation events

Looking back at theorem 1, one can imagine potentially handling larger information revelation events by integrating the dY terms. That turns out to be approximately possible, in a certain sense. Dytso, Poor, and Shamai (2022) prove the following.

Proposition 3 *[Dytso, Poor, and Shamai (2022), equation (52)] For a random variable x_t and a signal $y_t \sim N\left(x_t, \sigma^2\right)$,*

$$\frac{d}{dy}\kappa_j\left(x_t \mid y_t = a\right) = \kappa_{j+1}\left(x_t \mid y = a\right) \tag{20}$$

$\kappa_j\left(x_t \mid y_t = a\right)$ denotes the jth posterior cumulant of x_t conditional on observing $y_t = a$ for some fixed $a \in \mathbb{R}$.

This result shows that the type of recursion in theorem 1 continues to hold for discrete revelation events – diffusive information coming in infinitesimal increments in continuous time is not necessary for the central results. At the same time, it shows that normality is important – we can drop continuity, but proposition 3 still requires normality.[17] That said, proposition 3 also shows why continuous time is useful here: it allows us to use prior cumulants when calculating sensitivities. Specifically, even with knowledge of all the prior

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

cumulants and the value of the signal, one cannot directly use (20) since it involves the *posterior* cumulants.

4.4 Is it possible to ignore the higher-order cumulants?

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

That also means that if any of the higher cumulants is *ever* nonzero, then the distribution is *permanently* non-normal (since a Gaussian update of a non-normal distribution always yields a non-normal posterior), and all of the higher cumulants vary over time according to the dynamics in theorem 1.

5 Illustrative calibration

This section presents a simple quantitative example. Its motivation is twofold. First, it helps understand the extent to which the qualitative predictions above map into quantitatively reasonable behavior. Second, it helps in understanding the extent to which layering incomplete information over a standard and simple specification for fundamentals enriches the model's predictions in ways that allow it fit the data well. That said, it is important to emphasize that the simulation results are just an *example*. Their failure to match the data on some dimension does not mean that there is no model with the sort of learning we have studied so far that would do better, just that the exact specification detailed in this section is (obviously) imperfect.

5.1 Model setup

We assume that cash flows follow a dynamic version of the exponential example from above. That is, d_t is compound Poisson-exponential, with

$$dd_t = \phi\lambda dt - Y_t dN_t \tag{21}$$

where N_t is a Poisson process with constant rate ϕ and Y_t is an exponential random variable with mean λ. The $\phi\lambda dt$ term is a normalization so that mean dividend growth is equal to zero. For simplicity, we assume that discount rates are constant, so that $x_t = d_t - cons$.

The only free parameters in the model are ϕ, λ, and σ_Y. When $\sigma_Y^2 = 0$, this is a simple disaster model with constant disaster risk.

5.2 Numerical solution

The analytic results in principle require tracking an infinite number of cumulants over time. To simulate, we simply discretize the model – i.e. treat d_t as discrete Markov chain – and then directly calculate the updates via Bayes' theorem. We use the analytic formulas above when calculating conditional moments, and the results are nearly identical when calculating them numerically (i.e. via quadrature).

5.3 Parameter selection

We obtained parameters through simple moment matching. The aim here is not to maximize the data likelihood but rather to ask whether this extremely simple and obviously misspecified model can quantitatively fit major features of returns. The moments used for fitting are discussed in appendix A.1.5. Ultimately what is important is the model's quantitative behavior given the parameter – this is not an estimation exercise.

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

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

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

Matching those moments as well as possible leads to the calibration $\{\phi, \lambda, \sigma_Y\} = \{0.00037, 0.43, 2.89\}$, where the time unit is taken to be a day. That value of ϕ implies that disasters occur on average once every 10.7 years. $\sigma_Y = 2.89$

5.4 Results

Tables 1 and 2 report unconditional moments for returns and the VIX in the model and the data. Across the 13 moments, the model broadly matches the data, missing on only the most extreme statistics. Looking at returns, it has similar volatility and skewness, but kurtosis is too small by half. Note that kurtosis is necessarily the most weakly estimated of the three and most strongly driven by outliers. For returns scaled by lagged volatility, the model generates somewhat less kurtosis and skewness than is observed in the data, possibly indicating that it is failing to capture some intraday dynamics.

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

Table 1: Daily return moments

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

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

Table 2: VIX moments

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

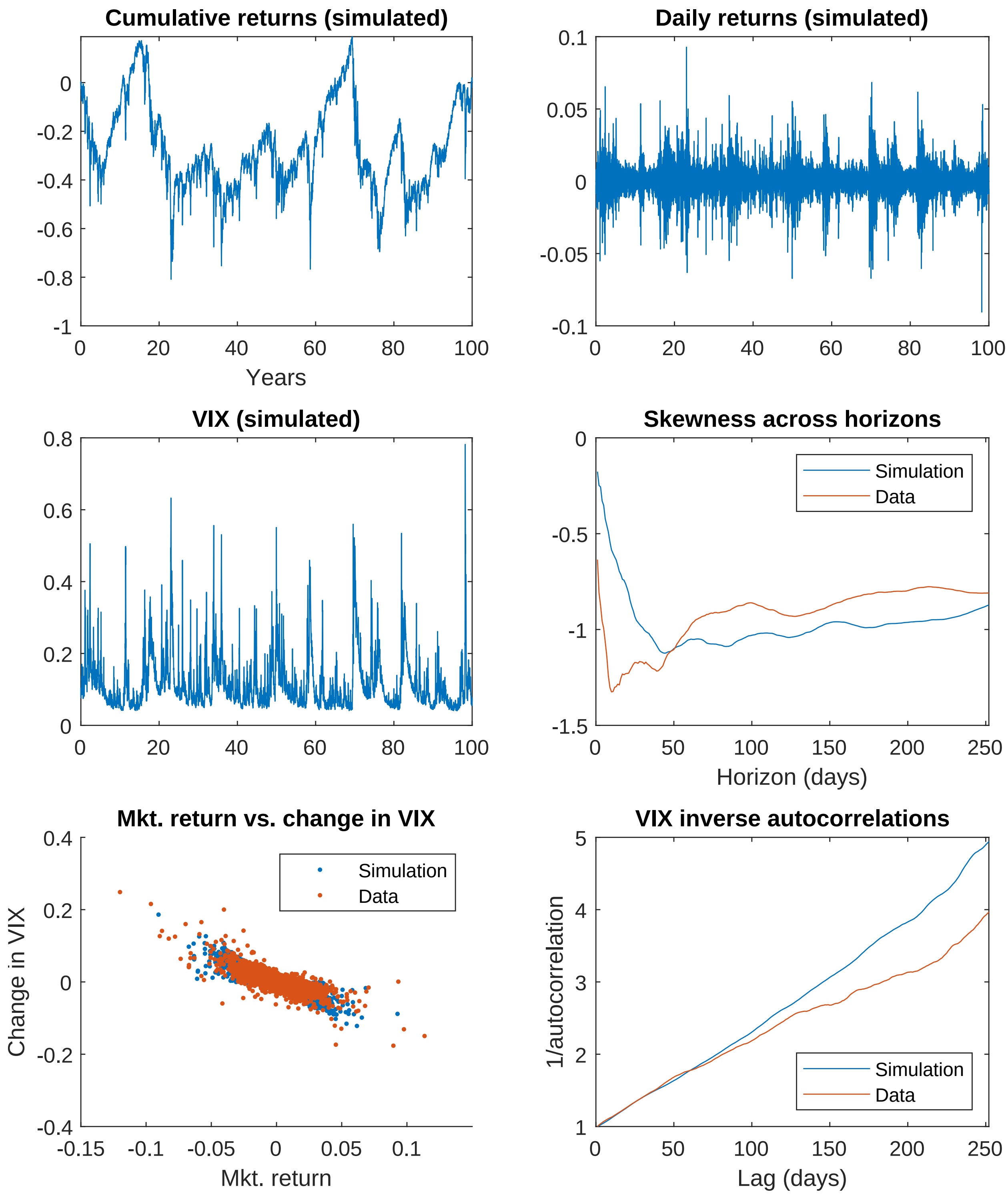
Note: The table reports moments of the VIX (level and daily changes), in the model and in the data.

The figure plots the behavior of asset prices in the simulated model. Panels (a), (b), and (c) are meant really just as an eyeball test – beyond the moments reported in tables 1 and 2, they show that the model generates returns and volatility that *appear*, both in levels and changes, highly similar to what is observed empirically. Panel (d) shows that the behavior of skewness across horizons is also similar, though it accumulates somewhat more slowly in the model.

Panels (e) and (f) report results for volatility dynamics. Panel (e) shows that the scatter plot of market returns against changes in the VIX is, in gross appearance, extremely similar in the model and data, while panel (f) shows that their autocorrelations – plotted here again in terms of the inverse (motivated by the analytic results for the microcosm model) – are nearly identical out to a year.

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

Figure 3: Simulations of motivating facts



Note: These are the same plots as those in figure 1, but for a 100-year simulation of the quantitative model.

Obviously what the model misses is risk premia – it has nothing to say at all about premia, whether on average or how they vary over time. The point of the exercise is to show that the joint dynamics that we observe for prices and volatility are a natural consequence of learning.

6 Estimated volatility dynamics and investor uncertainty

The analytic results in section 3 have specific implications for the dynamics of volatility and the leverage effect, and we explore them in this section. While the results in the previous section applied to one particular calibration, the results here are much more general tests of the model. We begin by estimating the dynamics of conditional volatility, then analyze volatility of volatility. The various coefficients are all functions of the parameter $\sigma_{Y,t}$. In some of the regressions we treat it as constant, and we also provide some evidence on whether that assumption is reasonable. Last, we show how the estimates can be used to obtain an estimate of $\kappa_{2,t}$ – agents' uncertainty about the fundamental value of the stock market, and we show that the estimate is similar in magnitude to survey evidence on cross-sectional disagreement.

6.1 Volatility dynamics

6.1.1 Regression setup

Combining equations (8), (6), and (10), we have

$$d\left[std\left(dp_t\right)\right] = \frac{1}{\sigma_{Y,t}} E_t\left[\left(dx_t\right)^2\right] + \frac{1}{\sigma_{Y,t}^2}\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right) dp_t - \frac{1}{\sigma_{Y,t}}\left[std\left(dp_t\right)\right]^2 dt \tag{22}$$

If x_t has independent increments and $\sigma_{Y,t}$ is constant – as in the quantitative model – then the first term is a constant.

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

We proxy for the second term with the product of the Cboe's SKEW index (properly transformed to correspond to a conditional skewness coefficient) multiplied by the CRSP total market return.[18] The $dt^{-1/2}$ term in parentheses in the middle term corresponds to the horizon at which skewness is calculated. We treat a unit of time as a day, so $dt = 21$, representing 21 trading days, which is the approximate horizon of the Cboe SKEW index (one calendar month). We include the factor 3, so that the regressor is $\left(3 \times skew_t^{Cboe} \times 21\right)$, where $skew_t^{Cboe}$ is the transformed Cboe SKEW index.

Finally, we proxy for the conditional volatility, $std\left(dp_t\right)$, with the Cboe's VIX index, rescaled consistent with the units for time. The time series are all daily. For the return dp_t we use the log return on the CRSP total market index. The actual regression, then, is the daily change in VIX_t on lagged implied skew times the market return (divided by 3) and VIX_{t-1}^2

There is a similar regression for the volatility of volatility. Again using (6) and (10) and combining now with (11), we have

$$std\left(dstd\left(dp_t\right)\right) = \frac{1}{\sigma_{Y,t}^2} \times \frac{\left|skew_{t\to t+\Delta t}\left(dp_t\right)\right|}{3} \times std\left(dp_t\right) \tag{23}$$

This says that the volatility of volatility is proportional to volatility itself times conditional skewness. We again proxy for volatility and skewness using the CBOE indexes. We proxy for $std\left(dstd\left(dp_t\right)\right)$ with the VVIX index, which is an implied volatility for the VIX itself.[19]

This regression now gives us a *third* estimate of σ_Y, and thus an additional testable prediction of the model. Additionally, the equation implies that the coefficient in the regression should be equal to zero, which is obviously also testable.

Finally, there is a fourth regression, related fairly closely to (22). In each month, we can estimate the leverage effect from a regression of changes in the VIX on market returns. According to equation (9), if we then regress the estimated leverage effect in month t on implied skewness in month t (divided by 3), the coefficient is again an estimate of $1/\sigma_Y^2$.

6.1.2 Results

Table 3 reports results of the regression implied by (22). The coefficients are highly statistically significant and have the expected sign. Under the model, if the various assumptions we made to derive the regression here are true, the coefficient on $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\left(dp_t\right)\frac{1}{3}\right)$

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

[19]Really, the VVIX measures the volatility of the log of the VIX. We therefore multiply the VVIX by the VIX to get the volatility of the VIX in its own units.

should be the square of the coefficient on VIX^2_{t-1} (proxying for $[std\,(dp_t)]^2$). The bottom row of the table tests that hypothesis. The t-statistic is -1.05, so the relative values of the coefficients are well within the range expected given statistical uncertainty.

Table 3: Volatility regressions

	(1)	(2)	(3)
VIX^2_{t-1}	-0.56	-0.68	-0.49
	[0.12]	[0.38]	[0.11]
$\frac{1}{3\sqrt{21}} skew_{t-1} dp_t$	0.46	0.46	0.22
	[0.01]	[0.01]	[0.02]
VIX_{t-1}		0.005	
		[0.010]	
dp_t			-0.039
			[0.004]
Constant			
R^2	0.62	0.62	0.65
Test of coefficients	-0.14		
	[0.14]		

Note: Daily regressions of first differences in the VIX. The skew is the Cboe Skew index, and dp is the log return on the CRSP total market index. Heteroskedasticity-robust standard errors are reported in brackets.

In the data, $[std\,(dp_t)]^2$ is obviously fairly strongly correlated with $std\,(dp_t)$ itself, so one question is which dominates – the model says it should be $[std\,(dp_t)]^2$. The second column of table 3 tests that proposition by including both in the regression. We find that $[std\,(dp_t)]^2$ does appear to dominate – its t-statistic is larger by a factor of 4 than that for $std\,(dp_t)$, and it remains marginally statistically significant (at the 10-percent level).

Similarly, we can ask whether it is $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\,(dp_t)\,\frac{1}{3}\right) dp_t$ that dominates or simply the return, dp_t. The third column of the table tests that by including both variables. In this case, dp_t turns out to have a very slightly higher t-statistic, but both variables remain individually highly significant. We thus have a less strong confirmation of the model's prediction here – dp_t should have been driven out, but it is not. On the other hand, it still does not drive out the correct variable, $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\,(dp_t)\,\frac{1}{3}\right) dp_t$. One possible explanation for this fact is that the Cboe SKEW index may be relatively noisy, since it is a higher-order moment, which would reduce the explanatory power of $\left(\Delta t^{-1/2} skew_{t\to t+\Delta t}\,(dp_t)\,\frac{1}{3}\right) dp_t$ (and also bias its coefficient toward zero).

Table 4 report results from the regression (23). The first column includes a constant –

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

Table 4: Volatility-of-volatility regressions

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

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

6.1.3 Estimates of the noise in investors' signals

It is possible to also obtain estimates of σ_Y on a rolling basis. One that is very direct is to just us (23), since that, in the absence of any error, gives a direct estimate of σ_Y in each period. However, obviously there is some noise, again both due to misspecification (the model is imperfect) and potential noise in asset prices. We therefore report results for five-year rolling windows. The coefficient in (23) corresponds to $1/\sigma_Y^2$.

Figure 4 plots the rolling estimates. Over the available sample, it appears relatively stable, moving between about 0.3 and 0.4.

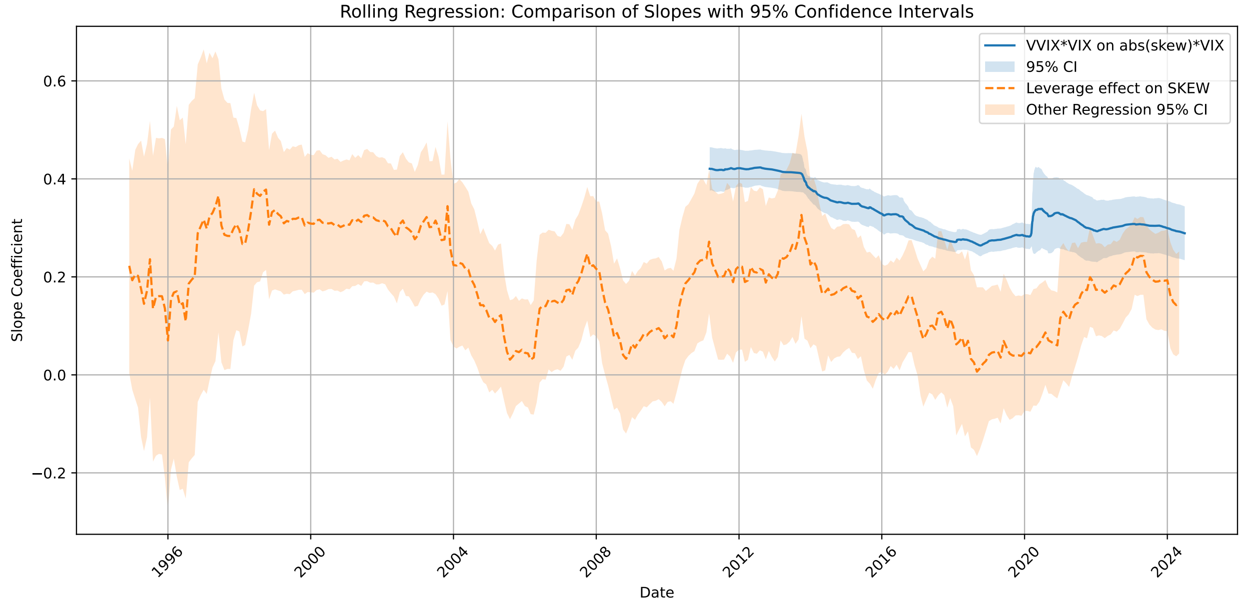
We can also get a rolling estimate based on the leverage effect. Specifically, combining equations (9) and (10), we have

$$\frac{\text{cov}\left(dp_t, d\left[std\left(dp_t\right)dt^{-1/2}\right]\right)}{\text{var}\left(dp_t\right)} = \frac{skew_{t\to t+\Delta t}\left(dp_t\right)}{3\sigma_{Y,t}^2} \tag{24}$$

The left-hand side is the coefficient from a regression of changes in volatility on changes in prices, which we can run each month. The ratio of that to $skew_{t\to t+\Delta t}\left(dp_t\right)$ at the beginning of the month is again an estimate of $1/\sigma_Y^2$.

This second time series for $1/\sigma_Y^2$ has somewhat lower values, but is broadly consistent

Figure 4: Rolling five-year estimate of $1/\sigma^2_{Y,t}$



Note: The blue line is the estimate of $1/\sigma^2_Y$ from equation (23) over a five-year rolling window. The orange line is based on the leverage effect regression. Confidence bands are calculated using Newey–West with 63 lags (three months of trading days).

with the first. Again, they are both fairly stable over time – though in the case of the leverage effect the confidence bands are significantly wider. That does not necessary rule out the possibility that it has some variation at higher frequencies. Distinguishing that from measurement error is subtle, though, and we leave it for future work.

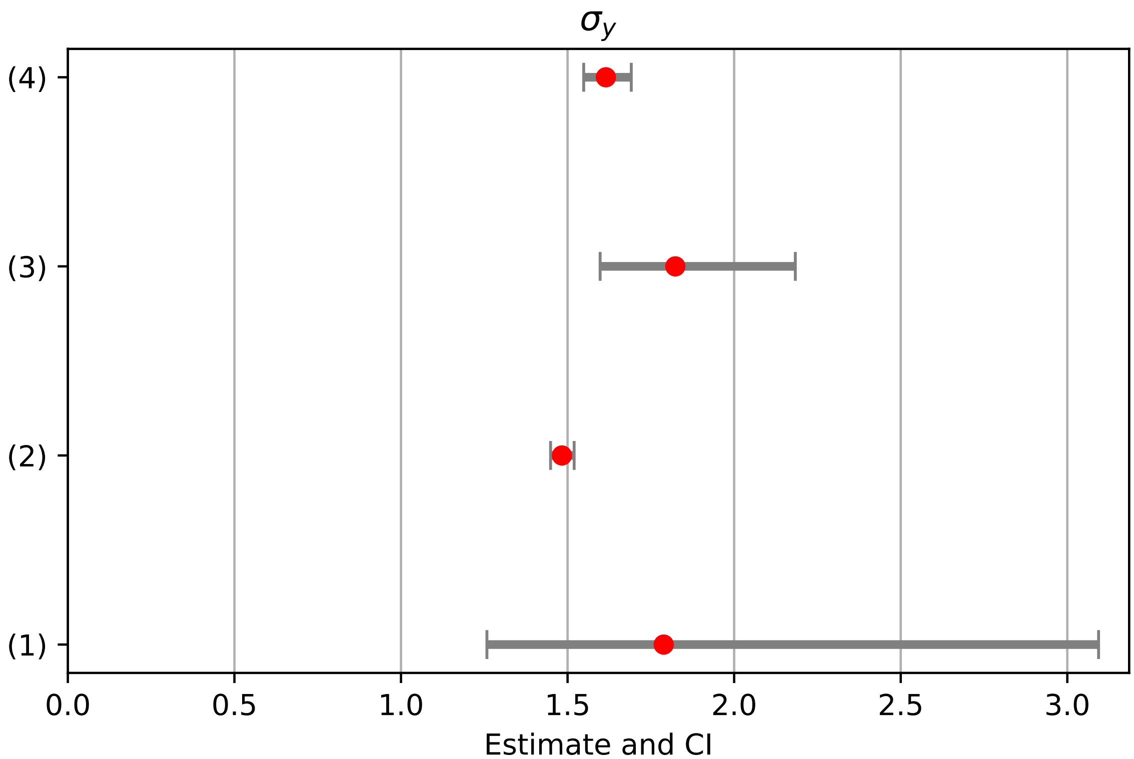
6.1.4 Comparing estimates of σ_Y

At this point we have four different estimates of σ_Y. Figure 6.1.4 compares them.

Estimate (1) is from the coefficient on VIX_{t-1} in equation (22); (2) is from the coefficient on lagged implied skewness times market returns in the same regression; (3) is from the volatility-of-volatility regression (23); and (4) is from the regression of the leverage effect on lagged skewness (the full-sample version of the orange line in figure 4). The four estimates are all surprisingly consistent, ranging between about 1.5 and 1.8. Estimates (2)-(4) are all forms of the relationship between skewness, returns, and volatility. Estimate (1) is somewhat more independent, being based on the rate of mean reversion in volatility.

Again, part of what makes these estimates notable is that there are in a sense nonparametric: They do not require knowledge of the true dynamics of fundamentals. Additionally, the estimates are not completely inconsistent with the value of 2.89 used in the calibration

Figure 5: Comparison of estimates of σ_Y



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

(which was obtained from a moment-matching exercise). While the value is higher than those implied from skewness and the leverage effect (estimates (2)-(4)), it is within the confidence band for the estimate based on mean reversion in volatility.

6.1.5 Estimates of investors' uncertainty about fundamentals

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

$$std\left(dp_t\right) = \frac{\kappa_{2,t}}{\sigma_{Y,t}} dt^{1/2} \tag{25}$$

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

Recall that the scaling of the estimates is for a unit time interval being equal to a day. The observed historical daily standard deviation of stock returns is about 1 percent. If $\sigma_{Y,t}$ is between 1.26 and 3.10 (based on the coefficient on VIX_{t-1} in equation (22), which is the most conservative of the confidence intervals), that implies that agents' posterior standard deviation is between 11.2 and 17.6 percent. The ± 2 standard deviation range for fundamentals around the current price for the aggregate stock market is then between ± 22.4 and ± 35.2 percent.

Similarly, we can get an estimate of average skewness in beliefs. One-month return skewness is historically approximately -2.2 (based on the average of the SKEW index). Plugging that into (10) along with the estimates of κ_2 and σ_Y yields an estimate for the skewness of fundamentals between -0.29 and -1.13. In the time series, the estimate of conditional skewness of fundamentals is proportional to the conditional skewness of returns divided by the square root of the conditional standard deviation of returns.

We have not yet found a survey that directly measures investors' uncertainty about fundamentals (e.g. that asks them about probabilities that the fundamental value might fall in different ranges, as the *Survey of Consumer Expectations* and *Survey of Professional Forecasters* do for inflation and other variables). However, uncertainty is sometimes proxied for by disagreement, so a survey giving a cross-section of estimates of fundamental value would be one way to validate our estimate of average uncertainty.

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

The data runs from August, 1993 to July, 2024 and has 8,242 observations. In that sample, we estimate the cross-sectional standard deviation to be **17.0 percent**, which fits inside the confidence band for uncertainty from equation (26) of [11.2,17.6]. That said, if we used the narrower confidence bands from the other estimates of σ_Y, the implied uncertainty would be somewhat lower.

6.1.6 Summary

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

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

7 Conclusion

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

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

References

Axenie, Cristian, Oliver López-Corona, Michail A Makridis, Meisam Akbarzadeh, Matteo Saveriano, Alexandru Stancu, and Jeffrey West, "Antifragility in complex dynamical systems," *npj Complexity*, 2024, *1* (1), 12.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A.1 Proofs

A.1.1 Theorem 1

A.1.1.1 Assumptions

Assumption 1 *The net present value of cash flows, x_t, follows a Feller process with bounded and smooth functions in the domain of its extended infinitesimal generator.*[1]

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

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

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

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

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

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

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

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

A.1.1.2 Proof

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

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

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

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

Equation (8.2) is simply equation (2) of the paper in integral form. Since $f(x) \equiv \exp(i\omega x)$ is a smooth and bounded function, by Assumption 1, there exists a measurable function $\mathcal{G}_f$ such that, almost surely, $\int_0^t |\mathcal{G}_f(x_s)|ds < \infty$ for every t and $M_t \equiv f(x_t) - f(x_0) - \int_0^t \mathcal{G}_f(x_s)ds$ is a right-continuous martingale. Therefore, condition (8.1) is satisfied. Furthermore, since f is a bounded function, so is $\mathcal{G}_f$. Consequently, conditions (8.6) and (8.7) are also satisfied. Conditions (8.4), (8.5), (8.9), and the second part of condition (8.3) are satisfied by Assumption 3. Finally, condition (8.8) and the first part of condition (8.3) are satisfied by Assumption 2. Applying Theorem 8.1 and noting that the Brownian motion W_t is independent of x_t, we get

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

where

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

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

Or equivalently,

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

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

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

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

Theorem 1 *Let $\kappa_{k,t}$ denote the kth cumulant of the posterior distribution of x_t conditional on Y^t. Suppose the $n+1$th moment of the posterior distribution and the nth moment of x_t exist, and Assumptions 1–3 are satisfied. Then for every $k \leq n$,*

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

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

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

Proof. By Lemma 1 and Itô's lemma,

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

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

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

The key step is to differentiate equation (A.11). Differentiating the left-hand side with

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

respect to ω and applying the dominated convergence theorem we get

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

For the first term on the right-hand side of (A.11), since x_t has n moments, for any ω in a sufficiently small neighborhood of the origin,

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

Therefore,

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

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

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

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

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

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

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

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

For the second term on the right in (A.11), note that

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

Therefore,

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

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

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

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

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

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

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

Therefore, the induction base holds with

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

Now suppose the induction hypothesis holds for k. Then,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A.1.2 Local skewness derivation

$$E\left[\left(\frac{k_2}{\sigma^2}\varepsilon + \frac{k_3}{2\sigma^2}\varepsilon^2 - \frac{k_3}{2\sigma^2}\sigma^2\right)^2\right] = k_2^2\sigma^{-2}dt \tag{A.39}$$

$$E\left[\left(\frac{k_2}{\sigma^2}\varepsilon + \frac{k_3}{2\sigma^2}\varepsilon^2 - \frac{k_3}{2}dt\right)^3\right] = 3\frac{k_2^2}{\sigma^4}\frac{k_3}{2\sigma^2}3\sigma^4dt^2 - 3\frac{k_2^2}{\sigma^4}\frac{k_3}{2\sigma^2}\sigma^4dt^2 \tag{A.40}$$

$$= 3k_2^2k_3\sigma^{-2}dt^2 \tag{A.41}$$

$$skew = \frac{3k_2^2k_3\sigma^{-2}dt^2}{k_2^3\sigma^{-3}dt^{3/2}} = 3k_2^{-1}k_3\sigma^{-5}dt^{1/2} \tag{A.42}$$

A.1.3 volatility regression derivation

$$skew_t\left(dp_t\right) \approx 3\kappa_{3,t}\kappa_{2,t}^{-1}\sigma_{Y,t}dt^{1/2} \tag{A.43}$$

$$d\kappa_{2,t} = \frac{\kappa_{3,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{\kappa_{2,t}^{2}}{\sigma_{Y,t}^{2}}dt \tag{A.44}$$

$$std\left(dp_t\right) = \frac{\kappa_{2,t}}{\sigma_{Y,t}}dt^{1/2} \tag{A.45}$$

$$\left(skew_t\left(dp_t\right)/dt^{1/2}\right)\frac{1}{3}\sigma_{Y,t}^{-1} \approx \frac{\kappa_{3,t}}{\kappa_{2,t}} \tag{A.46}$$

****

$$\frac{d\kappa_{2,t}}{\sigma_{Y,t}} = \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}^{2}}{\sigma_{Y,t}^{2}}\right)dt \tag{A.47}$$

$$\frac{d\kappa_{2,t}}{\sigma_{Y,t}} = \frac{1}{\sigma_{Y,t}}\frac{\kappa_{3,t}}{\kappa_{2,t}}\frac{\kappa_{2,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}^{2}}{\sigma_{Y,t}^{2}}\right)dt \tag{A.48}$$

$$\frac{d\kappa_{2,t}}{\sigma_{Y,t}} = \frac{1}{\sigma_{Y,t}^{2}}\left(skew_t\left(dp_t\right)/dt^{1/2}\right)\frac{1}{3}\frac{\kappa_{2,t}}{\sigma_{Y,t}^{2}}\left(dY_t - E_t\left[x_t\right]dt\right) + E_t\left[d\left(x_t^2\right)\right] - \frac{1}{\sigma_{Y,t}}\left(\frac{\kappa_{2,t}^{2}}{\sigma_{Y,t}^{2}}\right)dt \tag{A.49}$$

****

A.1.4 Exponential model solution

A.1.5 Moments for parameter selection