Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence*

Ned Augenblick
University of California, Berkeley
Haas School of Business

Eben Lazarus
Harvard University
Department of Economics

July 2017

PRELIMINARY DRAFT
(Please do not distribute)

Abstract

We examine restrictions on the variation in asset prices — specifically, variation in risk-neutral beliefs expressed in option prices — that must hold in a broad class of rational-expectations equilibria. We derive an upper bound for intertemporal movement in risk-neutral beliefs under rational expectations, and we show that this implies a lower bound on the curvature of utility (or, more generally, the slope of the stochastic discount factor) required to rationalize the marginal investor's beliefs. These restrictions have empirical counterparts, and we conduct empirical tests using risk-neutral distributions measured from S&P 500 index option prices. We find empirically that very high utility curvature is required to rationalize the behavior of risk-neutral beliefs, and in some cases, no stochastic discount factor in the class we consider is capable of rationalizing these beliefs. Under the theory we develop, this can be understood as providing evidence of overreaction to new information, though we consider the possibility that alternative explanations or option-market idiosyncracies could be driving our results. We conclude by considering the statistical and macroeconomic correlates of our findings and possible theoretical channels underlying these results.

*Emails: ned@haas.berkeley.edu and elazarus@fas.harvard.edu. Eben Lazarus thanks John Campbell, Gabriel Chodorow-Reich, Emmanuel Farhi, Matthew Rabin, and Jim Stock for their outstanding guidance and support. We are also grateful to Joshua Abel, Arielle Bernhardt, Laura Blattner, Krishna Dasaratha, William Diamond, Benjamin Enke, Xavier Gabaix, Mark Gertler, Ben Golub, Gita Gopinath, Niels Gormsen, Sam Hanson, David Laibson, Matteo Maggiori, Mikkel Plagborg-Møller, Simon Potter, Andrei Shleifer, Emil Siriwardane, Robert Solow, Jann Spiess, David Sraer, Stefanie Stantcheva, Adi Sunderam, Neil Thakral, Johan Walden, Chenzi Xu, and seminar participants at Harvard and UC Berkeley–Haas for advice, comments, and helpful discussions.

1. Introduction

The assumption that individuals have rational expectations (RE) over future outcomes is central to many models in macroeconomics and finance. A large body of previous literature has examined this assumption empirically, often in the context of asset markets. Asset prices in part reflect investor expectations over future outcomes, so these markets would seem to provide ideal testing grounds to discipline our understanding of such expectations.

In much of this past literature, researchers have argued that asset prices appear excessively volatile relative to measures of fundamental value under the RE benchmark, and therefore that expectations are overly volatile. But these empirical exercises face a difficult challenge: when observing volatile asset-price changes, how can we disentangle whether these arise due to changes in expectations rather than risk preferences? More simply and informally, when prices decrease, have expectations of future dividends decreased or has the marginal investor simply become more risk-averse (or both)? This past literature has accordingly appealed to restrictions on risk preferences or the data-generating process for asset returns in making the claim that their findings in fact reflect excess volatility in expectations themselves.[1]

We make progress on these questions by considering a general theoretical framework in which we show that certain restrictions on asset-price movements must hold in a broad class of rational-expectations equilibria. These bounds have empirical counterparts, and they accordingly give us direct information on the restrictiveness of the rational-expectations assumption in the data: for any level of observed volatility in the asset prices we consider, our bounds give a precise and conservative value for the curvature of utility — or, in asset-pricing terms, the slope of the stochastic discount factor across states — required to rationalize the data. We show further that there are values of observed volatility that cannot possibly be rationalized by any amount of utility curvature, allowing us to disentangle risk aversion from excess volatility in beliefs in these cases.

We focus our analysis on option prices, and we gain traction theoretically by translating option prices into so-called *risk-neutral beliefs* over the underlying asset's future price. Risk-neutral beliefs represent a probability distribution that would be equal to a hypothetical risk-neutral agent's true belief distribution about the future asset price, but these risk-neutral beliefs are in general distorted relative to the marginal investor's true (or *physical*) beliefs in the case that the investor is risk-averse.[2] The risk-neutral belief distribution can be calculated directly using option prices — options allow for bets over the future asset price, and thus the prices of these bets allow us to back out a probability distribution over the future asset price — so we treat risk-neutral beliefs as observable values, as is standard. These risk-neutral beliefs are then our main object of inter-

[1] As noted in LeRoy and LaCivita (1981), for example, the benchmark tests in Shiller (1979, 1981), and later in De Bondt and Thaler (1985), Campbell and Shiller (1987), and others, can generally be interpreted as joint tests of stationarity, constant discount rates, and rational expectations. We can accordingly interpret their null rejections as joint rejections of these three assumptions. See the end of this section for a full literature review.

[2] Intuitively, with risk aversion, the probability distribution we observe using option prices will overweight states in which the marginal investor has low wealth (e.g., when the underlying asset has a low return), since the investor will be willing to pay more than the actuarially fair value to insure against these high-marginal-utility states.

est, and our bounds show theoretically how their intertemporal behavior is restricted under the maintained assumption of rational expectations.

Economically, these theoretical results reflect the following intuition about belief updating under rational expectations: when individuals' beliefs about a future outcome change, they must on average be getting closer to certainty about the ultimate realization of the event. If not, then they are, loosely, overreacting to new information relative to the rational (Bayesian) benchmark.

For example, consider a hypothetical agent's beliefs over whether a single outcome will occur at some future date T. For simplicity, assume for now that this agent is risk-neutral over the outcome's realization. She believes ex ante, as of date 0, that the outcome has a 10% chance of occurring. Then we observe her beliefs moving to 90% as of date 1, to 10% on date 2, to 90%, to 10%, and so on until date T, at which point her beliefs resolve to either 0% (if the outcome does not occur) or 100% (if it does). Observing one such stream of beliefs, we might simply conclude that she received extreme and alternating signals on each successive day, forcing her to reverse her beliefs dramatically on a day-to-day basis. But if we were to observe such a pattern repeatedly, we might instead conclude that she is systematically overreacting to new information; her beliefs are mean-reverting (or anti-persistent) in a predictable manner, which violates the martingale property of beliefs under Bayes' rule, as can be shown formally. Further, this matches the intuition from the previous paragraph: the fact that her beliefs are mean-reverting is equivalent to the fact that her beliefs are never moving closer to certainty — measured in terms of their distance from 0% or 100% — despite their large day-to-day changes.

The intuition in this example is present as well in Augenblick and Rabin (2017), who use a framework similar to ours to prove restrictions on beliefs in the case that we can directly observe those beliefs (as in the above example). When considering beliefs over future asset prices as expressed in options, this is again equivalent to the case in which the marginal investor is risk-neutral. Our task becomes considerably more difficult when the assumption of risk neutrality is dropped. In this case, we observe only risk-neutral beliefs rather than true beliefs, and these risk-neutral beliefs need not follow a martingale given their distortion relative to the true beliefs (which do follow a martingale under RE). The bulk of our theoretical work is thus in extending the logic from the preceding example to this more general case, and showing how risk-neutral belief movement can be similarly restricted under RE.

The key insight we exploit to obtain our main results in this general case is that unlike the underlying asset price or expected returns, beliefs are *bounded*: even with large risk aversion, risk-neutral beliefs can only be distorted a finite distance from the true underlying beliefs, since risk-neutral beliefs lie between 0% and 100% by definition. Using this insight, we obtain our formal notion of excess volatility by showing that intertemporal movement in risk-neutral beliefs is bounded under rational expectations for any given level of risk aversion. This bound thus provides our mapping between movement in asset prices (or risk-neutral beliefs) and the degree of risk aversion required under RE. This mapping arises naturally in this asset-pricing context: we prove our results in terms of risk-neutral beliefs (because they are observable), and the deviation

of these risk-neutral beliefs from the true underlying beliefs depends on the degree of risk aversion of the marginal investor.

Aside from the maintained assumptions of rational expectations and no arbitrage in asset markets, our main theoretical results require only one general assumption on the stochastic discount factor (SDF), the random variable that determine's an asset's ex-ante price by discounting the asset's random future cash flows. In particular, we assume that the realization of the SDF depends only on observable variables and does not depend on the path of state variables realized between a given trading date and the option expiration date. We refer to this assumption as *conditional transition independence (CTI)*, and this assumption is met in many common macro-finance models. (We later extend our theoretical results to incorporate cases in which this assumption is violated.)

We then take our bounds to the data. We construct test statistics implied by our theoretical results using daily data on S&P 500 index option prices obtained from OptionMetrics. Our empirical estimation thus gives us information on the degree of risk aversion (or the slope of the SDF) required to rationalize the observed volatility in the risk-neutral beliefs implied by options over the value of the market index.

Conducting this estimation, we find that very high risk aversion is needed to rationalize the data, and in many cases our point estimates indicate that *no* amount of risk aversion is capable of rationalizing the data. Intuitively, the marginal investor's beliefs over the future price of the market index are overly volatile relative to the RE benchmark; equivalently, these beliefs (and the associated risk-neutral beliefs) have a tendency to mean-revert.[3]

Given that we conduct our estimation using index options data, we must also consider whether idiosyncracies specific to this market could be responsible for some of our empirical findings.[4] For example, if bid-ask bounce induces spurious variation in measured risk-neutral beliefs, this could upwardly bias our estimate of the excess movement in those beliefs. We construct our benchmark empirical tests with such issues in mind — for example, we use daily data rather than higher-frequency data to avoid intraday bid-ask bounce — but we also consider whether related market-specific issues could nonetheless affect our results and attempt to account for them explicitly in robustness tests, which weaken our results slightly but still indicate that very high risk aversion

[3]These results suggest that many leading frameworks capable of explaining medium-to-low-frequency variation in asset prices have difficulty rationalizing medium-to-high-frequency variation in beliefs without very high risk aversion.

[4]In this case, one might still consider our findings to be indicative of some apparent inefficiency, but specific to the options market rather than with respect to macroeconomic beliefs more generally. There is indeed evidence of idiosyncracies in this market; for example, various put-selling and related strategies have high measured returns (Coval and Shumway, 2001; Bondarenko, 2003), and Jackwerth (2000) argues that a pricing kernel backed out from options data is nonmonotonic in the index return (though Barone-Adesi, Engle, and Mancini, 2008, argue otherwise). Santa-Clara and Saretto (2009) suggest that margin requirements account for much of the apparent mispricing, while Bollen and Whaley (2004) and Gârleanu, Pedersen, and Poteshman (2009) argue for the importance of demand pressures. However, Bates (2008), Broadie, Chernov, and Johannes (2009), Santa-Clara and Yan (2010), and Constantinides, Jackwerth, and Savov (2013) show evidence instead suggesting that the apparent mispricings are insignificant once peso-problem-type sampling uncertainty and disaster- or jump-risk premia are accounted for, and are therefore not specific to option markets: "Option and stock returns may remain puzzling relative to consumption and dividends, but there is little evidence for mispricing relative to the underlying stock index" (Broadie, Chernov, and Johannes, 2009). There is also some evidence (e.g., Buraschi and Jiltsov, 2006; Ziegler, 2007) for belief-based distortions, but this is similarly consistent with an interpretation based on factors not entirely specific to the option market. We return to this issue in Sections 4–6.

is needed to rationalize the data.

We further consider the features of the data that yield our empirical conclusions, leading to three main additional findings. First, we find that excess volatility is concentrated in trading periods relatively far from a given option maturity date: for the last two weeks of trading before expiration, the data can be rationalized with finite and reasonable values for risk aversion, while this is no longer true at longer horizons from expiration. Thus our findings arise largely because beliefs about events in the somewhat-distant future appear to react too strongly to new information. Second, re-conducting our estimation at different sampling frequencies, we find that the risk-aversion value required to rationalize the data is decreasing as one decreases the sampling frequency from daily to weekly to monthly. The month-to-month variation in risk-neutral beliefs is moderate enough to be explained by finite (but somewhat large) risk aversion, but this masks substantial volatility (and required risk aversion) at higher sampling frequencies. Third, conducting reduced-form regressions of our excess volatility measure on a range of macroeconomic statistics, we find that it has a strong positive relationship with measures of macroeconomic uncertainty and no relationship with measures of limits to arbitrage in asset markets.[5]

Finally, we consider the robustness of our results to violations of our assumption of conditional transition independence for the SDF. The habit-formation model of Campbell and Cochrane (1999) violates this assumption and is also capable of matching important asset-pricing moments in the data, so we consider a calibrated version of the model as an instructive example for how violations of conditional transition independence might affect our conclusions under the maintained null of RE. We solve and simulate the model numerically,[6] and two findings emerge from these simulations: (a) even given the violation of CTI, the simulated risk-neutral beliefs still exhibit substantially less variation than observed in empirical data; (b) perhaps more importantly, when we naïvely apply our theoretical bounds to these simulated beliefs to estimate the risk aversion required to rationalize the simulated data, the bounds are still conservative as estimates of the true underlying risk aversion inherent in the model. This indicates that our bounds still apply approximately under mild violations of CTI, and we formalize this in an additional theoretical result. We consider these results to be suggestive evidence that violations of this sort have difficulty accounting for the excess volatility in beliefs we observe in the data.

Relation to previous literature and interpretation of results. Our paper relates to and builds on multiple separate but related strands of previous literature. We follow Shiller (1979, 1981), and later, among others, LeRoy and Porter (1981), De Bondt and Thaler (1985), Campbell and Shiller (1987), and Stein (1989) in conducting empirical tests for excess volatility in asset prices relative to the RE benchmark.[7] This literature proposes multiple different tests of excess volatility; taking

[5] This may be considered additional evidence against the possibility that factors specific to the option market are the main driver of our results, but this regression evidence is admittedly only suggestive and reduced-form.

[6] For our purposes, solving the model numerically requires solving for the joint distribution of t-period-ahead ($t \in \{1, 2, \ldots, T\}$) realizations of the equity return and the SDF. Given that this is a high-dimensional object, we apply projection methods to solve for this distribution, and these numerical methods may be of interest in their own right.

[7] A related literature — e.g., Fama and French (1988a,b), Poterba and Summers (1988), and later Campbell and Thompson (2008) — presents evidence on the predictability of returns. As noted by Campbell and Shiller (1988a,b),

Shiller (1981) as a benchmark case, he considers the variance of an aggregate equity index price relative to the variance of the present value of the ex-post stream of future dividends discounted at a constant rate, and finds that the former is much greater than the latter. But as noted by LeRoy and LaCivita (1981), this test, and those that followed, are in fact joint tests of (a) stationarity of the relevant data-generating process (for prices, in Shiller's case); (b) constant discount rates; and (c) rational expectations. Kleidon (1986) and Marsh and Merton (1986) emphasize the importance of possible non-stationarity in accounting for these results; meanwhile, much of the modern asset-pricing literature rationalizes the observed price volatility by appealing to time variation in discount rates (see Campbell, 2003, Cochrane, 2011, and Ludvigson, 2013, for surveys).

We build on the excess-volatility literature by showing that even without imposing any restrictions on the structure of the data-generating process, and imposing only mild restrictions on the variation in discount rates, RE nonetheless restricts the admissible variation in option prices in an empirically testable way. There are two costs associated with this additional generality.

First, we must consider derivative prices (specifically, risk-neutral beliefs over index returns) rather than directly considering the behavior of the underlying index, as in much of this previous literature.[8] In this way, our work is complementary to recent work by Giglio and Kelly (2017), who document excess volatility for claims on equity and currency volatility, inflation swaps, commodity futures, and credit default swaps. They do so using a framework that, like ours, accounts in principle for discount-rate variation: they exploit the (approximately) linear factor structure of cash flows for those assets under the risk-neutral measure in many standard models (as in Chamberlain and Rothschild, 1983), and they show how this factor structure restricts the price variation in long-maturity claims relative to that of short-maturity claims and that this restriction is violated in the variance-ratio tests they conduct in the data. Aside from considering different asset classes — we consider index options, while their framework leads them to consider mostly swaps — we differ further in that our estimation procedure and our focus on the risk-neutral belief distribution allow us to consider the specific behavior of beliefs that leads to the excess volatility documented in both papers, which is less straightforward in the context of variance-ratio tests for prices.[9,10] But despite these differences, we consider the two approaches to be complementary, and both yield similar evidence.

Second, rather than considering a setting allowing for fully binary (rejection vs. non-rejection) empirical tests of benchmark models, our general framework instead allows for a mapping be-

excess price volatility implies return predictability, so these papers' findings are related to the findings in the excess-volatility literature.

[8] We have in fact obtained additional results that restrict the variation in the underlying price, but these are preliminary and not yet included in this draft.

[9] There are further technical differences between the Giglio and Kelly (2017) paper and ours. For example, their estimation procedures require the assumptions of either (a) stationarity in the term structure of prices *and* no measurement error in their factor model for short horizons, *or* (b) homoskedastic Gaussian forcing processes for the factor model, while we do not require such restrictions. Giglio and Kelly do, however, consider possible violations of those restrictions and argue that their findings are robust to such violations.

[10] Additional literature examining benchmark models using variance ratios, from which we differ, includes Lo and MacKinlay (1988), Cutler, Poterba, and Summers (1989), and Campbell (1991), and Cochrane (1992).

tween the observed asset-price variation and the risk aversion required to rationalize the data. (We do, however, show that there are cases in which even infinitely large risk aversion cannot rationalize this observed variation; we postpone a discussion of the interpretation of this result to Section 4.) Our results may thus appear similar in spirit to those of Mehra and Prescott (1985), and more generally Hansen and Jagannathan (1991), who find that the SDF must be highly volatile to rationalize the observed excess returns for risky assets. Our results differ from theirs in two respects. First, we obtain our mapping using the second moment (i.e., the variation) of observed returns, while they use the first moment (or average) of returns.[11] More importantly, the Hansen-Jagannathan results may in principle be explained by features of the data-generating process for consumption or returns rather than high risk aversion per se; for example, models of rare disasters (e.g., Rietz, 1988; Barro, 2006; Gabaix, 2012; Martin, 2013; Wachter, 2013) can generate sufficient SDF volatility to rationalize the observed equity premium without requiring high risk aversion. But this is importantly not the case for our results. We obtain a relationship between local changes in the risk-neutral belief distribution and local risk aversion (or the slope of the SDF) at those points of the distribution, so that our finding of large belief variation across the entire distribution of beliefs *cannot* be rationalized by features of the data-generating process for consumption or returns; instead, conditional on our identification assumptions, this finding is attributable either to very high risk aversion or a departure from the RE benchmark we consider.[12] Put more simply, if we observe highly variable risk-neutral beliefs over the event that the S&P's 90-day return will be between 8% and 10%, we know that this cannot be attributable to disasters that affect the left tail of the return distribution.

Our approach is, however, related to that of Hansen and Jagannathan (1991) at a somewhat higher level: we maintain the spirit of their general semi-parametric setting, as we use a sufficient-statistic-type approach to recover structural parameters from observable data. In this way one may also relate our work to the sufficient-statistics literature in other fields; for example, Chetty (2006) derives and estimates an upper bound on risk aversion using labor-supply behavior, and Chetty (2009) provides a longer survey of the literature in settings different from ours. Within the asset-pricing literature more specifically, see, for example, Alvarez and Jermann (2005), Backus, Chernov, and Zin (2014), and Martin (2017), among many others.

Briefly summarizing our relation to literature more closely related to the specifics of our empirical setting and findings, our results complement evidence on beliefs obtained from survey data rather than asset prices, as, for example, in Frankel and Froot (1987), Han (2008), Greenwood

[11] We could instead map between variation in returns and the volatility of SDF volatility (i.e., an object related to the heteroskedasticity of the SDF), but we consider the results from our mapping to be somewhat more intuitive than this alternative.

[12] Our framework admits an informal interpretation of such a rejection as arising due to overreaction to information, and we frequently refer to our findings in this manner. But we should note as well that the set of alternatives being tested against the RE benchmark we consider does in principle include models in which agents do not make this particular statistical error: models of ambiguity aversion (e.g., Gilboa and Schmeidler, 1989; Hansen and Sargent, 2001, 2008; Epstein and Schneider, 2010), or of heterogeneous beliefs with short-sale constraints (e.g., Harrison and Kreps, 1978; Geanakoplos, 2010; Simsek, 2013), for example, could account for our empirical findings. We postpone a full consideration of the possible theoretical underpinnings of our empirical results to the Conclusion.

and Shleifer (2014), Gennaioli, Ma, and Shleifer (2016), and Manski (2017), as well as the results of Augenblick and Rabin (2017) for settings in which beliefs are directly observable. Another set of related literature endeavors to measure physical beliefs (rather than risk-neutral beliefs) directly using options data, for the purpose of examining either expectations or preferences. Aït-Sahalia and Lo (2000) and Jackwerth (2000) assume that the conditional subjective (and objective) physical distribution over returns is equal to a function of the observed historical distribution (cf. Aït-Sahalia, 1996, Stutzer, 1996, for the unconditional distribution); Aït-Sahalia, Wang, and Yared (2001), Poteshman (2001), Pan (2002), and Rosenberg and Engle (2002) fit parametric models to prices and/or returns that allow their distributions to be time-varying; Bliss and Panigirtzoglou (2004) restrict the representative-investor utility function to be either power utility or constant-absolute-risk-aversion utility, and then estimate the single unknown parameter in either case by minimizing forecast errors for the distribution of returns constructed using the risk-neutral densities and the assumed utility function.[13] Ross (2015) assumes a Markov process for transitions between return states and a transition-independence assumption on the pricing kernel similar to but more restrictive than the one we use (see Borovička, Hansen, and Scheinkman, 2016, for a discussion), from which Perron-Frobenius theory allows him to back out a distribution of physical beliefs. Our approach differs from the above set of literature in that we need not measure physical beliefs at all or know the true data-generating process for returns to conduct our tests, so we accordingly require less structure.[14]

Organization. The remainder of the paper is organized as follows. Section 2 introduces our theoretical framework and the intuition for our results by considering a simple two-state example, first stated with respect to directly-observed beliefs and then extended to consider the effects of risk aversion. Section 3 then presents a general many-state framework in which an arbitrary stochastic discount factor prices assets, and our theoretical bounds in this general case are collected in Section 4. Section 5 describes the data we use to conduct our empirical test and presents our estimation strategy and main empirical results, while Section 6 conducts additional empirical tests to consider the statistical and macroeconomic correlates of these results and their robustness. Section 7 discusses possible theoretical channels for our results and concludes. Proofs are contained in Appendix A, and Appendix B contains additional technical detail from Sections 3–6.

[13] Bates (2003) and Garcia, Ghysels, and Renault (2010) provide surveys of related literature, and see also Backus, Chernov, and Martin (2011) for the application and comparison of multiple related methods to estimate beliefs over consumption disasters. And as discussed in Section 5, the implementation of our empirical tests builds as well on this literature and related econometrics literature: we use the results of Breeden and Litzenberger (1978) and the curve-smoothing methods of Bliss and Panigirtzoglou (2004) and Malz (2014) to measure risk-neutral probabilities using options data, and we use the block bootstrap (Kunsch, 1989) and methods from the partial-identification literature (Imbens and Manski, 2004; Tamer, 2010) for estimation and inference.

[14] Similar parametric concerns apply as well to the option-anomalies literature discussed in Footnote 4, and these concerns are in a sense isomorphic to the peso problem that Broadie, Chernov, and Johannes (2009) argue underlies the anomaly results: if, for example, one measures the physical distribution of returns using historical data (e.g., Jackwerth, 2000), then an absence of crashes in the sample will lead one to overestimate the ex-ante returns to a put-selling strategy and to incorrectly infer the shape of the pricing kernel over return states. Linn, Shive, and Shumway (2015) argue that this is an empirically relevant concern, and see also Weitzman (2007) for broader discussion.