Robust Identification of Investor Beliefs

Xiaohong Chen[*]   Lars Peter Hansen[†]   Peter G. Hansen[‡]

October 28, 2020

Abstract

This paper develops a new method informed by data and models to recover information about investor beliefs. Our approach uses information embedded in forward-looking asset prices in conjunction with asset pricing models. We step back from presuming rational expectations and entertain potential belief distortions bounded by a statistical measure of discrepancy. Additionally, our method allows for the direct use of sparse survey evidence to make these bounds more informative. Within our framework, market-implied beliefs may differ from those implied by rational expectations due to behavioral/psychological biases of investors, ambiguity aversion, or omitted permanent components to valuation. Formally, we represent evidence about investor beliefs using a novel nonlinear expectation function deduced using model-implied moment conditions and bounds on statistical divergence. We illustrate our method with a prototypical example from macro-finance using asset market data to infer belief restrictions for macroeconomic growth rates.

Keywords— subjective beliefs, asset pricing, intertemporal divergence, bounded rationality, large deviation theory

---

[*]Yale University. Email: xiaohong.chen@yale.edu, ORCID:0000-0003-1125-675X
[†]University of Chicago. Email: lhansen@uchicago.edu, ORCID:0000-0003-0035-8558
[‡]MIT Sloan. Email: pghansen@mit.edu, ORCID:0000-0001-6351-0540

Prices in asset markets reflect a combination of investor beliefs and their risk preferences. Researchers, as well as policymakers, look to asset market data as a barometer of public beliefs. Derivative claims prices potentially enrich what we can infer about conditional probability distributions of future events, but events of interest often entail components of macroeconomic uncertainty for which there will be a paucity of information along some dimensions. Moreover, since a central tenet of asset pricing is that investors must be compensated for exposure to macroeconomic shocks that are not diversifiable, beliefs about future macroeconomic performance are of paramount importance to understanding asset prices.

1   Introduction

To disentangle the contributions of risk aversion from beliefs, many empirical approaches in the last few decades have focused on models of investor preferences by assuming rational expectations. Using the implied moment conditions of the investor's portfolio choice problem in conjunction with this restriction gives a directly applicable and tractable approach for estimating and testing alternative model specifications. This approach, however, often leads to risk prices in some time periods that are attributed to an arguably extreme level of investor risk aversion, or a rejection of the model. Risk aversion and belief formulation are intertwined. Rational expectations as a model of belief formation is meant to be a simplifying approximation. It can serve as an elegant and powerful modeling choice when appropriate. In a complex environment, however, it can be challenging to make statistical inferences pertinent to forward looking decision making. In such settings, we find the presumption that model-dwelling investors and entrepreneurs know the true data generating process to be tenuous, and worthy of relaxation. We are not alone in this view.

Some researchers have explored mechanisms that could account for this evidence via a different channel, namely beliefs which differ from rational expectations. It is sometimes argued, but typically not justified formally, that these alternatives are small departures from rational expectations. These "belief distortions" relative to rational expectations alternatively could reflect the lack of investor confidence about the assignment of probabilities to future events. This has been modeled and captured formally as ambiguity aversion or concerns about model misspecification.

This paper proposes a formal methodology for analyzing models that imply conditional moment restrictions where the restrictions are presumed to hold under a distorted probability measure. It extends a previous econometrics literature that represents the statistical implications of asset pricing implications as conditional moment restrictions under rational expectations. Rational expectations on the part of individuals or enterprises can be motivated by a Law of Large Numbers approximation used to pin down the beliefs of these economic agents "inside the model." Once we relax the rational expectations, there are typically many choices of investor beliefs that satisfy the conditional moment restrictions. Rather than imposing a specific alternative to rational

expectations, we restrict the family of investor probabilities to satisfy discrepancy bounds, which gives us a way of relaxing rational expectations hypothesis based on pushing back from a Law of Large Numbers approximation. We then use the conditional moment restrictions along with statistical discrepancy bounds, to characterize families of probabilities that satisfy the conditional moment restrictions.

Our approach provides a version of "bounded rationality" when assessing empirical evidence. Not only are the bounds we deduce of direct interest, they also can be used as diagnostics for specific models of belief distortions. Additionally, we show how to include survey data on subjective beliefs. Such data are typically sparse and not sufficient to pin down full probabilistic characterizations of beliefs. Thus even with these data, we are lead to bound probabilities, even when certain features of the data may be known through direct evidence.

A common way to represent a probability distribution of a random vector is through how it assigns expectations to functions of that random vector. Since we have multiple probability distributions in play, we represent our bounds by building what is called a "nonlinear expectation" that minimizes expectations over members of the family of probability distortions that we identify. This gives us a formal way to characterize properties of probability distributions that are consistent with model-implied conditional moment restrictions. The choice of minimization is essentially a normalization as we bound expectations of functions of observable variables and well as the negative of such functions. Given the flexibility in our choice of what functions of observables we use when forming expectation bounds, our analysis provides a rich characterization of implications for the belief distortions.

While we use asset pricing applications for motivation, our analysis is more generally applicable to economic models with forward-looking agents. These agents may be groups of individuals making investment or portfolio choices when facing production or financial opportunities that are exposed to uncertainty in different ways. Alternatively they may be forward-looking enterprises, making decisions today that have important consequences for the future.

In summary, our methodology gives a way to extract information on investor beliefs from asset market and survey data pertinent for both external analysts and for policy makers who are looking for evidence to gauge private sector sentiments. In addition, our computations provide revealing diagnostics for model builders that embrace specific formulations of belief distortions as is common in the behavioral economics and finance literatures.

Literature Review

There is a long intellectual history exploring the impact of expectations on investment decisions. As was well-appreciated by economists such as Pigou (1927), Keynes (1936), and Hicks (1939), investment decisions are in part based on people's views of the future. Alternative approaches for modelling expectations of economic actors were suggested including static expectations, Metzler

(1941)'s extrapolative expectations, Nerlove (1958)'s adaptive expectations, or appeals to data on beliefs; but these approaches leave open how to proceed when using dynamic economic models to assess hypothetical policy interventions. A productive approach to this modeling challenge has been to add the hypothesis of rational expectations. Motivated by long histories of data, this hypothesis pins down beliefs by equating the expectations of agents inside the model to those implied by the data generating distribution. This approach to completing the specification of a stochastic equilibrium model was initiated by Muth (1961) and developed fully in Lucas (1972).

Recently there has been a renewed interest in alternative belief distortions within the asset pricing literature. See for example, Fuster et al. (2010), Hirshleifer et al. (2015), Barberis et al. (2015), Adam et al. (2016), and Bordalo et al. (2019). Relatedly, since survey evidence on investor beliefs is typically rather sparse and not able to produce entire predictive distributions, Piazzesi et al. (2020) and Bhandari et al. (2019) fit time series models to the observed beliefs that can be distinct from the actual data evolution. Our approach is different from these literatures, but complementary to them. We focus on the construction of the implied bounds for expectations for functions of the stochastic process of interest that could provide empirical targets of tests of parametric models of subjective beliefs fit to time series.

There is similarity in motivation and overlaps in the methods we use to the study of robust optimization, (see, for instance, Petersen et al. (2000) and Duchi et al. (2018)) and robust Markov chain modeling (see, for instance, Škulj (2009)). But our is aim different. While the robust optimization research features decision makers that confront multiple probabilities, our perspective is that of analyst seeking information about the beliefs of economic decision-makers from observed financial market data or survey data through the lens of a dynamic economic model.

Hansen and Singleton (1996) and Nagel and Singleton (2011) describe and implement econometric methods for confronting conditioning information under correct model specification under rational expectations, within a generalized method of moments framework. Gagliardini and Ronchetti (2020) and Antoine et al. (2018) give extensions of the measure of model misspecification proposed by Hansen and Jagannathan (1997) to accommodate conditioning information. Similarly, the models we consider are misspecified under rational expectations. This misspecification is induced as it is in precursors Hansen (2014) and Ghosh et al. (2017) by investor belief distortions.[1] Our innovation is to propose and justify a dynamic formulation with belief uncertainty that a) accommodates conditioning, b) uses the recursive structure of multi-period likelihoods to characterize families of beliefs that are consistent with alternative divergence thresholds.

Outline of the Paper

Section 2 introduces the framework we use for moment restrictions implied by an asset pricing model. Section 3 specifies the probabilistic environment that underlies our computations and

---

[1]These two papers abstract from the role of conditioning information.

gives a dynamic version of the divergence with a built-in recursive structure. Section 4 presents and justifies our recursive formulation of the functional equation used to compute the bounds along with some special cases that are of particular interest. This section provides a more complete characterization of the solution for the familiar relative entropy divergence and discusses the relation to results from Large Deviation theory. Finally, section 4 characterizes a nonlinear expectation as a way to represent the bounds on the subjective probabilities. Section 5 presents an empirical illustration of our methodology. Section 6 shows how to apply our approach to extract information about the one-period, risk neutral measure and the long-term counterpart without assuming the existence of data on a complete set of Arrow-Debreu securities. Both of these probability measures are of interest in their own right. Section 7 concludes.

2 Asset Pricing with Distorted Beliefs

In standard economic applications, moment conditions are justified via an assumption of rational expectations. This assumption equates population expectations with those used by economic agents inside the model. These expectations are therefore presumed to be revealed by the Law of Large Numbers applied to time series data.

Let $(\Omega, \mathfrak{G}, \mathsf{P})$ denote the underlying probability space and $\mathfrak{I} \subset \mathfrak{G}$ represent information available to investors. The original moment equations under rational expectations are of the form

$$\mathbb{E}\left[f(X, \theta) \mid \mathfrak{I}\right] = 0. \tag{1}$$

where the function, f, captures the parameter dependence, θ, of either the payoff or the stochastic discount factor along with a random vector, X, of variables observed by the econometrician and used to construct the payoffs, prices, and the stochastic discount factor.

A typical asset pricing example is as follows: Let R denote an n-dimensional vector of gross returns corresponding to payoffs on financial or physical assets over some investment horizon, let S denote the corresponding stochastic discount factor for this horizon, and let $\mathfrak{I}$ denote the investor information set. The stochastic nature of the stochastic discount factor captures the market compensations for exposure to uncertainty.

The underlying asset pricing equation is

$$\mathbb{E}[SR - \mathbf{1}_n | \mathfrak{I}] = 0$$

where $\mathbf{1}_n$ is an n-dimensional vector of ones. Both the stochastic discount factor and the return vector R may depend on unknown parameters, giving rise to (1). The vector of returns can be parameter dependent when the investment is in a physical asset with an unobserved return. While we posed this as an asset-pricing relation, restrictions of the same form can be derived from invest-

ment or asset demand equations with potentially differential exposures to uncertainty. Even in such models, there is a counterpart to stochastic discount factors that are perhaps different across agent types. More generally, these equations feature the forward-looking behavior of individuals or enterprises as they depend on the perceptions of the future.

2.1 Market beliefs

We allow for the beliefs that are revealed by the market to differ from the rational expectations beliefs implied by (infinite) histories of data. We represent what we call "market beliefs" by introducing a positive random variable M with a unit conditional expectation. Thus, we consider moment restrictions of the form:

$$\mathbb{E}\left[Mf(X,\theta) \mid \mathfrak{I}\right] = 0. \tag{2}$$

The random variable M provides a flexible change in the probability measure, and is sometimes referred to as a Radon-Nikodym derivative or a likelihood ratio. The dependence of M on random variables not in the information captured by $\mathfrak{I}$ defines a relative density that informs how rational expectations are altered by market beliefs.

By using M to represent a potential market belief, we require that any event that depends on the realization of X and has conditional probability measure zero under the rational expectation distribution will continue to have conditional probability zero under this change in distribution. We will, however, sometimes allow for M to be zero on events that have positive conditional probability under the original measure, at least as a possible limiting case. Such events would be a complete surprise to market participants.

The introduction of M into the analysis is seemingly an innocuous change in formulating the observable implications. But it has rather dramatic consequences for econometric analyses. Specifically, we consider estimation environments in which a researcher only uses data on a limited set of asset returns. With observations on a complete set of asset returns and a pre-specified stochastic discount factor, we could identify uniquely the belief distortion, M. Given our interest in macroeconomic risk compensation, we presume a more modest set of data is available to use as empirical inputs. As a consequence, even with a known stochastic discount factor, there may be an extensive family of beliefs that is consistent with the underlying pricing restrictions expressed as conditional moments. To elaborate, we suppose that equation (1) may not have solutions for any θ under rational expectations. Once we relax rational expectations by introducing M, equation (2) will in general be satisfied for an infinite dimensional set of possible M's for each value of θ. Thus, the parameter vector θ and the corresponding M fail to be point identified in a rather spectacular way. The set of M's associated with a given value of θ will be of particular interest to us.

Given our interest in the set of M's, we are led to deduce implied bounds on moments.

Consider, for instance, the relation:

$$\mathbb{E}\left(MSR \mid \mathfrak{I}\right) - \mathbf{1}_n = 0. \tag{3}$$

Then the proportional risk premia from the perspective of the altered probability is:

$$\log \mathbb{E}\left(MR \mid \mathfrak{I}\right) + \mathbf{1}_n \log \mathbb{E}\left(MS \mid \mathfrak{I}\right).$$

The first term is the logarithm altered expectation of R and the second term is the negative of the logarithm of the risk-free return. Our methods allow us to compare the rational expectations version of the risk compensations to bounds on these proportional compensations as implied by market data.

Two classes of asset pricing models that have received considerable attention provide motivation for our analysis. One class allows for subjective beliefs to differ from those implied by rational expectations because of "market psychology." Alternative models of expectations from behavioral finance imply alternative specifications of M. Another class includes models with investors that are ambiguity averse. Associated with many such models are belief specifications that emerge as altered probabilities encoded in asset prices. These distortions reflect some form of caution depending on modeling details. While both literatures derive counterparts to M, our methods put very modest structure on the beliefs beyond potentially small statistical departures from rational expectations and can provide revealing diagnostics for assessing models that impose specific distortions in expectations.

Since the form of our pricing equation applies to investment or asset demand equations, there is a direct extension of our analysis to the case in which there are distinct classes of economic agents with potentially different subjective beliefs or concerns about ambiguity aversion.

2.2 Incorporating survey evidence

When constructing our moment conditions, we could also include direct data on investor expectations to help inform the direction and magnitude of the subjective belief distortion from historical evidence. This would entail augmenting the moment conditions used to constrain beliefs to include the variable being forecasted minus the observed forecast all scaled by M.

Suppose we have data D on beliefs that reflect subjective expectations of $\widetilde{X}$. These data could include survey responses or analyst forecasts. We may include this in our analysis by imposing the conditional moment condition:

$$\mathbb{E}\left(M\widetilde{X} \mid \mathfrak{I}\right) = D. \tag{4}$$

In words, this restriction says that D is the best forecast of $\widetilde{X}$ under the subjective belief measure. Note that we can incorporate probabilistic forecasts into our framework by letting $\widetilde{X}$ be an

indicator function.[2]

Remark 2.1. *Time series of survey data are often shorter relative to data on returns or macroeconomic variables. This can be accommodated in our framework provided that there is sufficient time series variation for these data to add nontrivial incremental information to the analysis.*

3 Data generation and probability divergence

In this section we construct and use the dynamic counterpart to the statistical divergence measure. We focus initially on relative entropy as a measure of statistical divergence, a measure which frequently arises in the analysis of large deviations of stochastic processes with temporal dependence.[3] While we use relative entropy as a starting point, we go much further by extending the so-called ϕ divergence measures in ways that have a recursive structure that is very similar to that of relative entropy.

While the applications that interest us use Markov formulations, we relax this assumption in order to entertain non-Markov distortions. For this reason, we initially consider a stationary and ergodic formulation that nests stationary, ergodic Markov processes. This is the same environment used by Hansen (1985) to study bounds on statistical efficiency and Hansen and Richard (1987) to study testable implications of asset pricing models, both in the presence of conditional moment restrictions. The Hansen and Richard (1987) analysis imposes rational expectations in contrast to the analysis in this paper.

We start with a baseline probability triple $(\Omega, \mathfrak{G}, \mathsf{P})$ and a measurable one-to-one transformation $\mathbb{U}$ which is measure-preserving and ergodic under P. We use $\mathbb{U}$ to construct stochastic processes and filtrations.[4]

Let $\mathfrak{I}_0 \subset \mathfrak{G}$ depict information available at date zero. We use the transformation $\mathbb{U}$ to capture the information available at future dates via the recursion:

$$\mathfrak{I}_t = \left\{ \Lambda \in \mathfrak{G} : \mathbb{U}^{-1}\Lambda \in \mathfrak{I}_{t-1} \right\} = \left\{ \Lambda \in \mathfrak{G} : \mathbb{U}^{-t}\Lambda \in \mathfrak{I}_0 \right\}$$

We presume that information accumulates:

$$\mathfrak{I}_t \subset \mathfrak{I}_{t+1},$$

which in turn implies that $\{\mathfrak{I}_t : -\infty < t < +\infty\}$ is a filtration. Similarly, for any random variable

---

[2]See Manski (2018) and the published comments for an overview and discussion of the use of survey data in macroeconomics and Bordalo et al. (2020) for a probe into the impact of heterogeneity in the study of over-reaction.

[3]See, for instance, Donsker and Varadhan (1975) or Dupuis and Ellis (1997).

[4]A common specification of $\mathbb{U}$ is the shift transformation applied to the space of infinite sequences of vectors of real numbers.

B_0 that is $\mathfrak{I}_0$ measurable, we form B_t recursively

$$B_t(\omega) = B_{t-1}\left[\mathbb{U}(\omega)\right] = B_0\left[\mathbb{U}^t(\omega)\right].$$

Thus for each initial random vector B_0, there is a corresponding stochastic process $\{B_t : t \geqslant 0\}$ that is adapted to the filtration $\{\mathfrak{I}_t : t \geqslant 0\}$. Since $\mathbb{U}$ is measure-preserving, the process $\{B_t : t \geqslant 0\}$ is stationary.

3.1 Alternative probabilities

In what follows, we hold fixed the transformation $\mathbb{U}$ while considering alternative probability measures. Let Q denote an alternative probability distribution on $(\Omega, \mathfrak{G})$ that is measure-preserving and ergodic, and let Q_t be the restriction of Q to $\mathfrak{I}_t$. We consider only Q's for which there exists an $N_1 \geqslant 0$ that is $\mathfrak{I}_1$ measurable and satisfies:

$$\int B_1 d\mathsf{Q}_1 = \int \mathbb{E}\left(N_1 B_1 \mid \mathfrak{I}_0\right) d\mathsf{Q}_0 \tag{5}$$

for all bounded $\mathfrak{I}_1$ measurable random variables B_1. This N_1 necessarily satisfies $\mathbb{E}\left(N_1 \mid \mathfrak{I}_0\right) = 1$. When Q is distinct from P, there is a bounded random variable for which the Q and P distributions differ. Because both are ergodic, the Law of Large Numbers is applicable to both with distinct almost sure limit points. For the purposes of this analysis, the probability measure Q encodes the limits implied by the Law of Large Numbers.[5]

Form the product

$$M_T = \prod_{t=1}^{T} N_t. \tag{6}$$

Then under Q, the date-T conditional expectation of a bounded, $\mathfrak{I}_T$ random variable B_T is

$$\mathbb{E}\left(M_T B_T \mid \mathfrak{I}_0\right).$$

We think of M_T as a relative likelihood between two models over horizon T constructed recursively through the familiar likelihood factorization. We further restrict Q to imply stochastic stability:[6]

Definition 3.1. *We say that Q induces* **stochastic stability** *if for any B_0 that is $\mathfrak{I}_0$ measurable and satisfies $\int |B_0| d\mathsf{Q}_0 < \infty$,*

$$\lim_{T\to\infty} \mathbb{E}\left(M_T B_T \mid \mathfrak{I}_0\right) = \int B_0 d\mathsf{Q}_0.$$

---

[5]From a measure-theoretic perspective, Q and P cannot be equivalent. There exists events based on limits for which Q assigns probability one and P assigns probability zero and conversely.

[6]Stochastic stability as defined here is satisfied when the process is beta-mixing (or absolutely regular); see, e.g., Theorem 3.29 in Bradley (2007).

Definition 3.2. *The set $\mathcal{N}$ contains all N_1's for which there is a corresponding probability Q satisfying* (5) *and is stochastically stable.*

We presume that $N_1 = 1$ is in this set and hence P is stochastically stable.

3.2 Intertemporal divergences

First consider the Kublack-Leibler divergence. Represent the expected log-likelihood ratio as a sum of contributions for each date by using the recursive structure of a relative likelihood:

$$\mathbb{E}\left(M_T \log M_T \mid \mathfrak{I}_0\right) = \mathbb{E}\left(M_T \sum_{t=1}^{T} \log N_t \mid \mathfrak{I}_0\right) \geqslant 0.$$

Dividing by T and taking limits gives:

$$\begin{aligned}
\mathcal{R}(N_1) &= \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left(M_T \log M_T \mid \mathfrak{I}_0\right) \\
&= \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left(M_T \sum_{t=1}^{T} \log N_t \mid \mathfrak{I}_0\right) \\
&= \int \mathbb{E}\left(N_1 \log N_1 \mid \mathfrak{I}_0\right) d\mathsf{Q}_0,
\end{aligned}$$

which is the measure of relative entropy that we will use in our analysis. Notice that there is an explicit connection between N_1 and Q_0, which gives rise to a restriction that we impose when computing bounds.

Remark 3.3. *The relative entropy measure $\mathcal{R}(N_1)$ is the discrete-time analog to the relative entropy measure that is used in the Donsker-Varahadan large deviation theory applied to Markov processes.*[7]

Finite relative entropy restricts substantially the tail behavior of the probability distributions. For this reason we consider other divergences, but modified in order to exploit the recursive structure implied by the likelihood factorization. We use the conditional version of what is commonly called a ϕ divergence as an important building block:

$$\mathbb{E}\left[\phi(N_t) \mid \mathfrak{I}_{t-1}\right]$$

for a strictly convex function ϕ defined on $(0, \infty)$ with $\phi(1) = 0$.[8] There is an equivalent way to

---

[7]See Dupuis and Ellis (1997) and Varadhan (2008) and Dembo and Zeitouni (2009) (see chapter 3).
[8]For some ϕ divergences the domain can be extended to include zero.

represent this divergence that we use. Construct a function ψ such that

$$n\psi\left(\frac{1}{n}\right) = \phi(n).$$

It may be shown that ψ is also strictly convex with $\psi(1) = 0$. By design:

$$\mathbb{E}\left[\phi(N_t) \mid \mathfrak{I}_{t-1}\right] = \mathbb{E}\left[N_t\psi\left(\frac{1}{N_t}\right) \mid \mathfrak{I}_{t-1}\right]. \tag{7}$$

On the right-hand side, we use the conditional probability associated with N_t to compute expectations. The random variable $\frac{1}{N_t}$ is the Radon-Nikodym derivative of the baseline conditional probability with respect to the N_t conditional probability distribution. The function ψ in (7) plays a role analogous to ϕ once we change the probability we use in the expectation.

We use ψ to represent an intertemporal divergence measure:

$$
\begin{aligned}
\mathcal{R}(N_1) &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left(M_{t-1}\mathbb{E}\left[\phi\left(N_t\right) \mid \mathfrak{I}_{t-1}\right] \mid \mathfrak{I}_0\right) \\
&= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[M_t\psi\left(\frac{1}{N_t}\right) \mid \mathfrak{I}_0\right] \\
&= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[M_T\sum_{t=1}^{T}\psi\left(\frac{1}{N_t}\right) \mid \mathfrak{I}_0\right]
\end{aligned}
\tag{8}
$$

The limiting version of this measure as implied by the Law of Large Numbers for stationary, ergodic processes is:

$$\int \mathbb{E}\left[N_1\psi\left(\frac{1}{N_1}\right) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0 = \int \mathbb{E}\left[\phi(N_1) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0.$$

Notice that we use the conditional version of what is commonly called a ϕ divergence measure averaged using the altered stationary probability Q. This coincides with the relative entropy divergence when $\phi(n) = n\log n$. The supplemental information (SI) appendix A shows that this intertemporal divergence extension of ϕ divergence is convex.

Alternatively, there has been considerable interest in members of the Wasserstein class of divergences, including in the study of robust Markov chains and machine learning. Wasserstein divergences could also be used as conditional divergences that are averaged using the altered stationary probability Q.[9]

Remark 3.4. *Empirical likelihood methods (see Qin and Lawless (1994) and Owen (2001)) use*

---

[9]For example, see Eckstein (2019). Eckstein (2019) deduces a Laplace principle for Large Deviation theory and other approximations with an overlapping class of intertemporal divergences. While he poses a problem that is mathematically related, his formulation does not nest ours.

a divergence measure $\phi(n) = -\log n$ for iid data. The corresponding intertemporal divergence measure is

$$\mathbb{E}\left(-\log N_1\right). \tag{9}$$

Note that this measure uses the original baseline probability measure implied by P_0 and not Q_0 to average over the conditioning information. Were we to average using Q_0 in the divergence, then our analysis would apply. Elsewhere, we and others have argued that this strictly decreasing ϕ is not well suited for detecting departures from baseline probabilities restricted by moment conditions.

Remark 3.5. *Using positive random variables, M_T, to depict alternative probabilities for date-T events imposes absolute continuity (conditioned on date zero information). As we already noted, this same absolute continuity will not be true over the infinite future, however. Our division by T when constructing intertemporal divergence (8) for the altered probability to have different limits under the Law of Large Numbers.*

Remark 3.6. *Our analysis assumes that the underlying processes are stationary under the alternative Q's. However, the results we obtain will still apply under certain classes of nonstationarity processes. While we restrict Q to be measure-preserving and ergodic, any probability measure that is absolutely continuous with respect to Q will obey the same Laws of Large Numbers even though it may not be measure-preserving. Such measures may also imply the same intertemporal divergence measure (8). Moreover, specific examples that could be of interest are probabilities associated with Markov processes that eventually "escape" from their dependence on the initialization. While we do not explore such cases formally, these observation suggests that the bounds that we compute could be justified under an even broader set of probability measures.*

4   Moment bounds

We next present the recursive approach we use for computing probability bounds. To represent these bounds, we entertain a rich family of functions g and compute sharp lower bounds on the expectation $g(X_1)$. As a special case, $g(X_1)$ could be indicator of an event with its expectation being the probability of that event. We will be interested in other choices of g in our empirical illustration. We compute an upper bound on $g(X_1)$ by finding the negative of the lower bound on the expectation $-g(X_1)$. Formally, we are interested in solving

$$\inf_{N_1 \in \mathcal{N}} \int \mathbb{E}\left(N_1 g(X_1) \Big| \mathfrak{I}_0\right) d\mathsf{Q}_0$$

subject to the constraints

$$\mathcal{R}(N_1) \leqslant \kappa$$

$$\mathbb{E}[N_1 f(X_1)|\mathfrak{I}_0] = 0.$$

To compute a bound for a given choice of g, we will borrow an idea from the robust control literature. See, for instance, Petersen et al. (2000) and Hansen and Sargent (2001). We will initially solve a problem with a discrepancy penalty indexed by a parameter $\xi > 0$ and show how to solve a problem given ξ. We will then treat ξ as a Lagrange multiplier and trace out the implied discrepancies for each such ξ. In this way, ξ may be chosen to enforce the constraint $\mathcal{R}(N_1) \leqslant \kappa$. For notational simplicity we suppress the parameter dependence in the moments of $f(X_t)$ and $g(X_t)$.

In much of what follows, we aim to solve:

Problem 4.1.

$$
\begin{aligned}
\mu^* &= \inf_{N_1 \in \mathcal{N}} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left(M_T \sum_{t=1}^{T} \left[g(X_t) + \xi \psi\left(\frac{1}{N_t}\right) \right] \Big| \mathfrak{I}_0 \right) \\
&= \inf_{N_1 \in \mathcal{N}} \int \mathbb{E}\left(N_1 \left[g(X_1) + \xi \psi\left(\frac{1}{N_1}\right) \right] \Big| \mathfrak{I}_0 \right) d\mathbb{Q}_0
\end{aligned}
$$

subject to $\mathbb{E}\left[N_1 f(X_1) \mid \mathfrak{I}_0 \right] = 0.$

It suffices to optimize over N_1 because this choice determines the N_t for all $t \geqslant 1$ through the transformation $\mathbb{U}^{t-1}$ and M_T as constructed in (6).

4.1 Martingale construction

While we solve problem 4.1 via recursive methods, as a precursor to this, we use a convenient martingale construction used in connection with the objective function. Suppose that for $N_1 \in \mathcal{N}$, we find a random variable v_0 such that

$$\mathbb{E}\left(N_1 \left[g(X_1) + \xi \psi\left(\frac{1}{N_1}\right) - \mu + v_1 \right] \Big| \mathfrak{I}_0 \right) - v_0 = 0 \tag{10}$$

where

$$\mu = \int \mathbb{E}\left(N_1 \left[g(X_1) + \xi \psi\left(\frac{1}{N_1}\right) \right] \Big| \mathfrak{I}_0 \right) d\mathbb{Q}_0,$$

and where $v_1(\omega) = v_0[\mathbb{U}(\omega)]$ and v_0 is $\mathfrak{I}_0$ measurable. Then iterating on equation (10), it follows that

$$
\mathbb{E}\left(M_T \sum_{t=1}^{T} \left[g(X_t) + \xi \psi\left(\frac{1}{N_t}\right) - \mu \right] \Big| \mathfrak{I}_0 \right)
$$
$$
+ \mathbb{E}\left(M_T v_T \mid \mathfrak{I}_0 \right) - v_0 = 0.
$$

12

In fact,

$$\sum_{t=1}^{T}\left[g(X_t) + \xi\psi\left(\frac{1}{N_t}\right)\right] - T\mu + v_T - v_0$$

is a martingale under the probability measure $\mathbb{Q}$ with expectation zero. Moreover,

$$v_0 - \int v_0 d\mathbb{Q}_0 \tag{11}$$

$$= \lim_{T\to\infty} \mathbb{E}\left(M_T \sum_{t=1}^{T}\left[g(X_t) + \xi\psi\left(\frac{1}{N_t}\right) - \mu\right] \mid \mathfrak{I}_0\right)$$

provided that the almost sure limit on the right-hand side is finite. The random variable v_o is in fact only well defined up to a constant translation. Our recursive approach will simultaneously solve for v_0 used in the martingale construction and μ computed at the minimizing solution for N_1.

4.2 Recursive formulation

We use a rearranged version of equation (10) when posing the recursive formulation of the optimization problem.

Problem 4.2. *Find a pair (μ, v) that satisfies*

$$\mu = \inf_{N_1\in\mathcal{N}} \mathbb{E}\left(N_1\left[g(X_1) + \xi\psi\left(\frac{1}{N_1}\right) + v_1\right] \mid \mathfrak{I}_0\right) - v_0$$

subject to the constraint:

$$\mathbb{E}\left[N_1 f(X_1) \mid \mathfrak{I}_0\right] = 0$$

where $v_1(\omega) = v_0[\mathbb{U}(\omega)]$ and v_0 is $\mathfrak{I}_0$ measurable and μ is a finite number. This optimization problem determines the constant μ and the random variable v_0 up to a translation by a constant.

This problem is recognizable as a fixed point problem in (μ, v_0) captured by the relation between v_1 and v_0. While we posed this problem in terms of date zero and date one, given our presumed stationary data generation the problem could equivalently be stated in terms of date t and date $t+1$ for $t > 0$. The minimization over N_1 can be solved using convex duality methods familiar from the analysis of ϕ divergence measures. We let (μ^*, v_0^*) denote the solution to problem 4.2, and N_1^* be the corresponding minimizer. The following objects are of interest from this problem:

- the moment bound: $\int \mathbb{E}\left[N_1^* g(X_1) | \mathfrak{I}_0\right] d\mathbb{Q}_0^*$;

- the corresponding conditional moment: $\mathbb{E}\left[N_1^* g(X_1)|\mathfrak{I}_0\right]$;

- the implied divergence:

$$\int \mathbb{E}\left[\phi(N_1^*) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0^* = \int \mathbb{E}\left[N_1^*\psi\left(\frac{1}{N_1^*}\right) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0^*$$

where N_1^* solves problem 4.2 and Q_0^* is an implied stationary distribution.[10]

There are three features of problem 4.2 that require further comment. First, the minimization problem includes "continuation-value" adjustments depicted by v_0 and its next period counterpart v_1 along with the numerical value μ. We include these adjustments to account for the fact that the choice of N_1 has implications for future time periods. Second, limit (11) is not finite for all measure-preserving and ergodic probabilities measures Q. For the minimum to be attained, we presume that this limit is finite at the minimizing solution. Third, $\xi\mathbb{E}\left(N_1\psi\left(\frac{1}{N_1}\right) \mid \mathfrak{I}_0\right)$ acts as a per-period divergence penalty, but we may equivalently think of ξ as a Lagrange multiplier and subsequently maximize over ξ. Thus, we use ξ to index alternative problems that penalize the increment to the intertemporal divergence. The value μ^* of this objective depends on ξ, leading us to write $\mu^*(\xi)$. To impose a specific divergence constraint κ, we solve

$$\sup_{\xi>0} \mu^*(\xi) - \xi\kappa \tag{12}$$

$$= \sup_{\xi>0} \int \mathbb{E}\left(N_1^*\left[g(X_1) + \xi\psi\left(\frac{1}{N_1^*}\right) - \xi\kappa\right] \mid \mathfrak{I}_0\right) d\mathsf{Q}_0^*.$$

Alternatively, we can back out the implied κ for each ξ by computing the derivative $\frac{d\mu^*}{d\xi}$ or by computing directly the divergence associated with N_1^*. To determine the ξ sensitivity, many versions of the optimization problem could be solved in parallel using convex duality and dynamic programming methods. The computed bound $\mu^*(\xi^*)$ is for the unconditional expectation of $g(X_1)$, although we find the conditional expectation, $\mathbb{E}\left[N_1^* g(X_1) \mid \mathfrak{I}_0\right]$ that is a central part of the calculation, to be of interest in its own right. Finally, $\lim_{\xi\to\infty} \mu^*(\xi)/\xi$ reveals the minimum possible divergence subject to the conditional moment restrictions. This limit gives a lower bound on magnitude of κ used in our analysis. It can be computed directly by solving a counterpart to Problem 4.2.

4.3 Nonlinear expectation

Consider now the set $\mathcal{B}$ of bounded Borel measurable functions g to be evaluated at alternative realizations of the random vector X_1. Given a divergence bound κ, we construct a mapping $\mathbb{K}$

---

[10]In contrast to our analysis, Ghosh and Roussellet (2020) incorporate conditioning information to identify a subjective belief by minimizing the divergence given in (9) under P,

from functions g in $\mathcal{B}$ into the real line that assigns the computed bound. Formally, define:

$$\mathbb{K}(g) = \int \mathbb{E}\left[N_1^* g(X_1)|\mathfrak{I}_0\right] d\mathbb{Q}_0^*$$

where the right-hand side is computed using the value of ξ that solves (12). This mapping can be thought of as a "nonlinear expectation," as formalized in the following proposition.

Proposition 4.3. *The mapping $\mathbb{K} : \mathcal{B} \to \mathbb{R}$ given by $\int \mathbb{E}\left[N_1^* g(X_1)|\mathfrak{I}_0\right] d\mathbb{Q}_0^*$ implied by problem 4.2 for each $g \in \mathcal{B}$ has the following properties[11]:*

(i) if $g_2 \geqslant g_1$, then $\mathbb{K}(g_2) \geqslant \mathbb{K}(g_1)$;

(ii) if g is constant, then $\mathbb{K}(g) = g$;

(iii) $\mathbb{K}(\mathsf{r}g) = \mathsf{r}\mathbb{K}(g), \quad$ for a scalar $\mathsf{r} \geqslant 0$;

(iv) $\mathbb{K}(g_1) + \mathbb{K}(g_2) \leqslant \mathbb{K}(g_1 + g_2)$.

All four properties follow from the definition of $\mathbb{K}$. Property iv includes an inequality instead of an equality because we compute by solving a minimization problem, and the N_1's that solve this problem can differ depending on g.

Remark 4.4. *While $\mathbb{K}(g)$ gives a* lower *bound on the expectation of $g(X)$, by replacing g with $-g$, we construct an* upper *bound on the expectation of $g(X)$. The upper bound will be given by $-\mathbb{K}(-g)$. The interval*

$$[\mathbb{K}(g), -\mathbb{K}(-g)]$$

captures the set of possible values for the distorted expectation of $g(X)$ consistent with divergence less than or equal to κ.

Remark 4.5. *In our statement of problem 4.2, we suppressed the dependence of the function f on an unknown parameter vector θ in a parameter space Θ. But many applications will necessarily include parameter uncertainty. Provided that we can compute the solution to this problem quickly using duality, we can assess parameter sensitivity by perhaps solving this problem many times in parallel. Bounds that are robust to parameter sensitivity could be obtained by minimizing over the set Θ or over a family of probability distributions over Θ.*

Remark 4.6. *For some applications, it is of interest to bound ratios of expectations of functions g_1 and g_2 of X_1. Expectations conditioned on discrete events are such ratios and proportional risk compensations are logarithms of such ratios. As we elaborate in SI appendix E, bounds of ratios can be computed by first bounding expectations of $g_1(X_1) - \zeta g_2(X_1)$ for alternative choices of the real number ζ and then searching over ζ for the smallest ratio.*

---

[11]The first two of these properties are taken to be the definition of a nonlinear expectation by Peng (2004). Properties *(iii)* and *(iv)* are referred to as "positive homogeneity" and "superadditivity".

4.4 Dual problem

We show that the dual problem for the relative entropy divergence is equivalent to a principal eigenvalue problem. This gives a representation of the distorted measures that underlies the bounds and a revealing link to large deviation theory.

By a direct application of duality for $\phi(n) = n \log n$,

$$\mu + v_0$$
$$= \max_{\lambda_0} -\xi \log \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X_1) + \lambda_0 \cdot f(X_1) - \frac{1}{\xi} v_1 \right] \mid \mathfrak{I}_0 \right)$$

where the random vector λ_0 is restricted to be $\mathfrak{I}_0$ measurable and is the vector of Lagrange multipliers for the conditional moment restriction. See SI appendix C for a more complete development of the primal and the dual problems. Let $\epsilon = \exp\left(-\frac{\mu}{\xi}\right)$ and $e_0 = \exp\left(-\frac{v_0}{\xi}\right)$. Then an equivalent statement of the dual problem is:

$$\epsilon = \min_{\lambda_0} \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X_1) + \lambda_0 \cdot f(X_1) \right] \left(\frac{e_1}{e_0} \right) \mid \mathfrak{I}_0 \right).$$

In this optimization problem λ_0 is again restricted to be a $\mathfrak{I}_0$ measurable random vector and e_0 is restricted to be positive as is the real number ϵ.

When the state space is not discrete, this eigenvalue problem can have multiple solutions. While there could be multiple solutions to this eigenvalue problem, the next result identifies the eigenvalue of interest.

Lemma 4.7. *When there are multiple positive eigenvalue solutions for a given λ_0, at most one of them induces a probability measure that is stochastically stable.*

See SI appendix B for a proof.[12]

Proposition 4.8. *Problem 4.2 with $\phi(n) = n \log n$ can be solved by finding the answer to*

$$\epsilon = \min_{\lambda_0} \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X_1) + \lambda_0 \cdot f(X_1) \right] \left(\frac{e_1}{e_0} \right) \mid \mathfrak{I}_0 \right)$$

where

$$\mu = -\xi \log \epsilon$$
$$v_0 = -\xi \log e_0.$$

---

[12]Relatedly, Hansen and Scheinkman (2009), prove a counterpart to this result for continuous-time specifications in a Markovian environment. Moreover, Qin and Linetsky (2020) extend the Hansen and Scheinkman (2009) analysis by, among other things, relaxing the Markov assumption.

In this optimization problem, the random vector λ_0 is restricted to be a $\mathfrak{I}_0$ measurable random vector, the random variable e_0 is restricted to be $\mathfrak{I}_0$ measurable and positive, and $e_1(\omega) = e_0[\mathbb{U}(\omega)]$ with probability one. The real number ϵ is positive. The implied solution for the probability distortion is:

$$N_1^* = \frac{\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0^* \cdot f(X_1)\right] e_1^*}{\epsilon^* e_0^*}$$

where λ_0^* is the optimizing choice for λ_0 and (ϵ^*, e_0^*) are selected so that the resulting Q^* induces stochastically stability. The conditional expectation implied by the bound is

$$\mathbb{E}\left[N_1^* g(X_1) \mid \mathfrak{I}_0\right],$$

which in turn implies a bound on the unconditional expectation equal to

$$\int \mathbb{E}\left[N_1^* g(X_1) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0^*.$$

The implied relative entropy is

$$\int \mathbb{E}\left(N_1^* \log N_1^* \mid \mathfrak{I}_0\right) d\mathsf{Q}_0^*.$$

Remark 4.9. *Our characterization is reminiscent of the results from large deviation theory for Markov processes. As in our analysis, large deviation theory studies an undiscounted limiting problem. See, for instance, Dupuis and Ellis (1997) and Varadhan (2008) for valuable treatises on large deviation theory.*[13]

A more substantive link to large deviations helps us interpret the relative entropy bounds that we input into our analysis. When using the empirical probability to detect potential departures from the baseline model, there is typically a positive probability that the empirical distribution mistakenly detects a departure. For a fixed criterion, the probability of this mistake becomes increasingly small as the sample size gets large with a well defined rate characterized by large deviation theory. Under some additional regularity conditions, remarkably, the decay rate can be made to be arbitrarily close to the minimum relative entropy bound that we compute. Moreover, this theory computes excursions, represented probabilistically, that make the decay rate as small as possible. See SI appendix D for an elaboration.

While we draw on insights from large deviation theory, our ultimate aim is quite different from that theory. Nevertheless, we find it revealing to compute both relative entropies and related Chernoff entropy as described, for instance, in Newman and Stuck (1979) and Anderson et al. (2003) as part of a sensitivity analysis.

---

[13]In particular, the analysis in Chapters 7 and 8 in Dupuis and Ellis (1997) features discrete-time Markov specifications and large sample approximation in the formulation of a Laplace principle.

4.5 Markov specification

To proceed in a tractable way, we impose Markovian restrictions on the underlying data generating processes. Specifically we presume that $\{X_t : t \geqslant 0\}$ is a time invariant function of Markov process, appropriately restricted.

Assumption 4.10. $\{(X_t, Z_t) : t = 0, 1, ...\}$ *is a first-order Markov process for which the joint distribution of* (X_{t+1}, Z_{t+1}) *conditioned on* (X_t, Z_t) *depends only on* Z_t.

Given this assumption, the $\{Z_t\}$ process by itself is a first-order Markov process. We view both X_t and Z_t as observable. The triangular structure for the dynamic evolution allows us to use a more sparse representation of the conditioning information. The alternative probabilities that we explore are not restricted to be Markov, but the solution to the minimization problem will be, for reasons that are familiar from dynamic programming. With the Markov specification, we solve the recursion

Problem 4.11. *Find a pair* (μ, v) *that solves*

$$
\begin{aligned}
&\mu + v(z) \\
&= \min_{N_1 \in \mathcal{N}} \mathbb{E}\left(N_1\left[g(X_1) + \xi\psi\left(\frac{1}{N_1}\right) + v(Z_1)\right] \mid Z_0 = z\right)
\end{aligned}
$$

subject to the constraint:

$$
\mathbb{E}\left[N_1 f(X_1) \mid Z_0 = z\right] = 0
$$

This optimization problem determines the constant μ *and the random variable* v *up to a translation by a constant.*

While the primal problem "imposed" stochastic stability, it suffices to verify the stability of the process that we obtain as our candidate solution. Since it is Markovian, this restriction is satisfied when the process $\{Z_t\}$ is aperiodic and Harris recurrent.

5 Illustration

We illustrate these methods using a familiar asset pricing model with recursive utility investors as in Kreps and Porteus (1978) and Epstein and Zin (1991). While much of the asset pricing literature appeals to an arguably large risk aversion in conjunction with rational expectations when confronting data, we constrain risk aversion and instead explore belief distortions as an alternative expectation. We follow Campbell (1993) by allowing for market segmentation and

avoid the direct use of consumption data. We then ask what implications asset market data have for predicted consumption growth rates.

Much of the macro asset pricing literature imposes risk aversion that is arguably large at least for some states of the macroeconomy. Epstein et al. (2014) and Dillenberger et al. (2020) give alternative rationale for substantially restricting the risk aversion coefficient. While we do not view as a settled issue what precise bounds should be imposed on risk aversion, we assume a unit risk aversion in order to feature the role of belief distortions when confronting asset pricing evidence. Our choice of unity is admittedly for convenience. As we will see, this choice leads to some particularly simple asset pricing implications.

Let R^w denote a presumed observable return on wealth. As noted by Epstein and Zin (1991), the one-period stochastic discount factor under rational expectations is the reciprocal of the gross return on wealth when risk aversion is unity. Thus, under distorted beliefs represented by M,

$$S = M(R^w)^{-1}$$

where S is the one-period stochastic discount factor under rational expectations. We use this setup for our illustration.

As Epstein and Zin (1991) note, the consumption Euler equation for an investor implies[14]:

$$\mathbb{E}\left[M \log R^w \mid \mathfrak{I}\right] = -\log \beta + (1-\rho)\mathbb{E}\left[M \log G \mid \mathfrak{I}\right]$$

where G is the ratio of consumption growth over two adjacent time periods. The parameter β is the subjective discount factor, and ρ is the reciprocal of the intertemporal elasticity of substitution. By deducing bounds on the left-hand side, we may infer bounds on ρ times the market expectation of consumption growth of equity market participants expressed in logarithms. Recursive utility preferences are specified in terms of continuation values that determine the rankings of prospective consumption processes. As a rough approximation, when $\rho < 1$ the wealth is positively related to the continuation value, where both are relative to current consumption. Conversely, they are negatively related when $\rho > 1$. Thus for this model of investor preferences, whether ρ is larger or smaller than one impacts how we interpret the evidence based on conditioning information.

In our illustration, we draw on the literature that suggests returns can be predicted from dividend-price ratios. While there have been debates on how fragile this evidence is, we step aside from that discourse and take the predictability evidence on face value to illustrate our method. Given our direct use of dividend-price measures, we purposefully choose a very coarse conditioning of information and split the dividend price ratios into three bins using the three empirical terciles. We take the dividend-price terciles to be a three-state Markov process. Dividend-price ratios are

---

[14]See equations (17) and (18) of Epstein and Zin (1991) except that we allow for more general expectations.

known to be persistent, and this will be evident in our calculations.[15]

We implement our approach using quarterly data from 1954-2016. We use the return on CRSP value-weighted index to proxy forthe return on wealth. For asset returns, we use the return on a 3-month treasury bill, and the three Fama-French factor excess returns. We impose moment conditions for each return implied by equation (3), each scaled by three indicator functions for the terciles of the dividend-price ratio, giving a total of 12 moment conditions. All returns are converted from nominal to real returns using the deflator for nondurables consumption obtained from the BLS. We then apply the methods described in section 4 to bound functions of the return on wealth as measured by the value-weighted return.

In figure 1, we report the bounds on the beliefs about the expected log return, which under the assumption of unitary risk aversion coefficient are approximately proportional to the consumption growth rate belief when the subjective discount factor β is very close to one. The conditional expectation of log returns and the unconditional counterpart are all lower than their empirical counterparts. This observation follows by comparing the •'s with the boxes where the top and bottom of the boxes are the upper and lower bounds with a relative entropy constraint imposed at a magnitude that is twenty percent higher than the minimum. The minimum relative entropy rate implies a half-life of about 24 quarters for reducing the probability by fifty percent of mistakenly rejecting the rational expectations. Increasing this by twenty percent, reduces the half-life by the same percentage to about 20 quarters.[16] While our choice of a twenty percent increase is a bit arbitrary and used for illustration purposes, it is straightforward to compute bounds with other choices of divergence thresholds. Across alternative applications, a choice of twenty percent, independent of magnitude of the minimal possible divergence would be hard to defend. One nice aspect of relative entropy is that there is an explicit statistical interpretation that we find to be revealing, thus the magnitude of the divergence has meaning.

Interestingly, it is when we condition on the low value of the dividend-price ratio we find the box with the largest height (biggest difference between the upper and lower bounds). Also, the bounds on the unconditional distorted expectations are very similar to those we found for the low dividend-price ratio. As a robustness check, we repeated these calculations for a quadratic conditional divergence and found only very modest differences. See SI appendix F.

Not only are conditional means distorted, but so are the transition probabilities as reported in table 1. While the implied stationary probabilities are fairly evenly distributed over the three

---

[15]As an alternative starting point, we could use the regime probabilities from Markov switching models of Maheu and McCurdy (2000) as possible states along with the implied return distributions.

[16]The Chernoff entropy at the minimum is .0096 with a half-life of 72 quarters. The Chernoff measure is motivated by a common decay rate imposed on type I and type II errors of testing one model against another and is expected to be considerably smaller. We computed it using the approach described in Newman and Stuck (1979) for Markov processes. While symmetric, this measure is less tractable to implement and not included in the family of recursive divergences that we describe. We use it merely to provide, *ex post*, additional information about the magnitude of the bound.
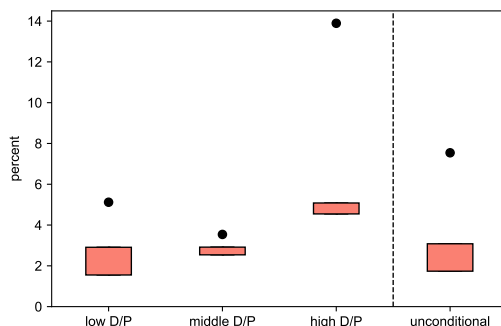
Figure 1: Expected log market return. The •'s are empirical averages and the boxes give the imputed bounds when we inflated the minimum relative entropy by 20%. The minimum relative entropy is .0284 with a half-life of 24.4 quarters.

dividend-states, essentially by construction, the minimal entropy probabilities down-weight substantially the high dividend-price ratio state and up-weight the low dividend-price state. The high dividend-price state, in particular, has a very small stationary probability under the minimum distorted stationary distribution. Consistent with this, the transition probabilities into this state are lower under the distortion and they are higher for exiting this state. The opposite happens for transitions in and out of the low dividend-price state. Thus, a hypothetical process that behaves in accordance with the minimum entropy distorted Markov transition matrix is likely to spend substantially more time in the low expected log return state and much less time in the high expected log-return state. When we increase the relative entropy bound by twenty percent, the implied distorted transition matrices are quite similar to the implied transition matrix recovered by the minimizing relative entropy and depart from the empirical transition matrix in comparable ways.

	empirical	min entropy
transition matrix	$\begin{pmatrix} .96 & .04 & 0 \\ .05 & .88 & .07 \\ 0 & .08 & .92 \end{pmatrix}$	$\begin{pmatrix} .98 & .02 & 0 \\ .08 & .88 & .04 \\ 0 & .17 & .83 \end{pmatrix}$
stationary probabilities	$\begin{bmatrix} .42 & .31 & .27 \end{bmatrix}$	$\begin{bmatrix} .76 & .20 & .04 \end{bmatrix}$

Table 1: Empirical and distorted transition probabilities.

An alternative approach would be to solve static versions of our analysis for each of the three different specifications of the conditioning information. Under this approach, there would be no reason for distorting the transition probabilities as the dynamical evolution of the conditioning
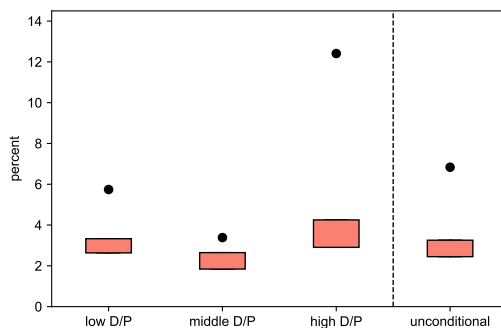
Figure 2: Proportional risk compensations computed as $\log \mathbb{E} R^w - \log \mathbb{E} R^f$ scaled to an annualized percent. The •'s are the empirical averages and the boxes give the imputed bounds when we inflated the minimum relative entropy by 20%.

information is ignored. Not surprisingly, this approach does lead to notable differences in bounds. The minimized entropy over the alternative configurations of the conditioning information using the undistorted transition probabilities is .047 in comparison to the much smaller .028 that we found using our method. We view belief distortions in the transition probabilities are of particular interest in behavioral models and in models with ambiguity aversion and see this as a virtue over methods that analyze the conditional problems separately.

As we mentioned at the outset of this section, there is a substantial asset pricing literature that studies time-varying risk compensation, often appealing to high values of risk aversion. We illustrate how belief distortions can imitate large risk compensations. Thus, consider the bounds on the implied risk compensations when we restrict the risk aversion parameter to be one. We report proportional risk premium using the ex-post real return on treasury bills, R^f, as our riskless benchmark in figure 2. To construct these results, we compute bounds on $\log \mathbb{E} R^w - \log \mathbb{E} R^f$ by extending the approach of section 4 as described in SI appendix E. Restricting investor risk aversion allows for belief distortions to capture the fluctuating empirical compensations for exposure to uncertainty.

Finally, since our empirical method looks at other information from two other Fama-French excess returns, we also report bounds on other risk compensations that we included in our analysis. We convert the excess returns into returns by adding the gross returns on bonds. We report these findings in SI appendix E. Again the risk compensations are greatly reduced relative to the empirical counterpart, and in most instances they are less sensitive to conditioning information.

Putting aside the empirical debate on return predictability, we see two possible conclusions from these results. One possibility is that the statistical divergence (measured as relative entropy) for the distortions is high enough to challenge a "bounded rationality" view of the recursive utility model with a unitary risk aversion. The other possibility is that this divergence is defensible, in

which case our dynamic implementation reveals the most statistically plausible distortions on the evolution of the dividend-price ratios. It remains a judgement call as to when the resulting statistical bounds we find here are implausible. Researchers that embrace rational expectations do not consider belief distortions, while behavioral finance researchers seldom consider the implied statistical divergence of their modeled beliefs. Neither practice uses tools for assessing statistical approximation as we have done in this paper. We view these computations as providing valuable empirical complement to the assessment of specific models of belief distortions or ambiguity aversion. Moreover, we have noted how survey evidence can be included within our framework, and we view this as a potentially valuable extension of the illustration presented here.

6 Bounding other probabilities

While we have motivated our analysis by a method for extracting expectation bounds for subjective beliefs or restricted "worst-case" beliefs that support valuation under ambiguity aversion. These same methods provide expectation bounds for two other probability measures that are of interest in asset valuation. These measures are the one period risk-neutral probabilities and the long-term forward probabilities. While surveys cease to provide information about these probabilities, even sparsely observed asset values, as we assume here, are revealing. We now comment on how to apply our method for both of these applications.

6.1 Risk-neutral measure

Under risk-neutral pricing, the reciprocal of the gross one-period riskless return acts as a stochastic discount factor. Thus, in this case:

$$MS = M(R^f)^{-1}.$$

Stutzer (1995, 1996) and Avellaneda (1998) target the M's with the *smallest* relative entropy divergence to use in pricing derivative claims. We map this type of problem into our analysis by viewing the empirically relevant distribution as the "correct distribution" and the risk neutral transformation as a way to correct for model misspecification. As in Stutzer (1996) and Avellaneda (1998), the measures of particular interest to us are the ones with a small divergence, although we explore more probabilities than just the M with the minimal divergence. While not our primary motivation, the methods we develop in this paper allow the user to obtain robust bounds on risk neutral expectations of macroeconomic variables that incorporate information embedded in asset prices.

6.2 Long-term forward measure

A substantively distinct, but mathematically related, literature studies the martingale decomposition of the stochastic discount factor. This decomposition expresses the stochastic discount factor as the product of a martingale component and a transitory component. The martingale component can be interpreted as a change of probability measures that imposes risk neutrality in valuation over long investment horizons. Kazemi (1992) and Alvarez and Jermann (2005) show that the reciprocal of the gross holding period return on a long-term bond is the stochastic discount factor net of a martingale component.

Since the work of Alvarez and Jermann (2005), the martingale component is referred to as the permanent component to the cumulative stochastic discount factor process. In providing a more formal mathematical characterization, Hansen and Scheinkman (2009) and Qin and Linetsky (2020) find it more revealing to appeal to probabilistic characterization of this component. As emphasized by Borovička et al. (2016), the probability measure associated with the martingale absorbs long-term risk adjustments for stochastically growing cash flows. This probability measure is the risk neutral, forward measure that captures long-term risk-return tradeoffs.

Many structural models of asset pricing have stochastic discount factor processes with martingale components that dominate risk prices over long investment horizons. These components can reflect permanent shocks to the macroeconomy or forward-looking components to valuation. These components are present when investors have recursive utility preferences in which the intertemporal composition of risk matters or when they are averse to ambiguity in assigning probabilities to future events.

To relate this to our analysis, suppose that this martingale component is missing from the model specification. In such circumstances, Kazemi (1992) justifies the use of the reciprocal of the gross holding period return on a long-term bond, R^h, as the stochastic discount factor: $S = (R^h)^{-1}$. When there is a martingale component, Alvarez and Jermann (2005) advocated bounding its magnitude by, in effect, using this return reciprocal as a misspecified stochastic discount factor. For this application, we use

$$S = M(R^h)^{-1}$$

as the stochastic discount factor where $M > 0$ has conditional expectation equal to one and thus induces a change of a probability measure that absorbs long-term risk adjustments. Our method applied to this problem complements and extends those of Alvarez and Jermann (2005) and Bakshi and Chabi-Yo (2012) by providing bounds on expectations implied by this measure.

Remark 6.1. *Ross (2015) gives conditions under which investor beliefs can be uniquely identified from asset prices. This identification hinges on a complete panel of state prices being available to the econometrician, as well as an implicit restriction that the stochastic discount factor process*

does not contain a martingale component. This latter restriction is violated in many standard asset pricing models (see Borovička et al. (2016) for a discussion). More generally this identification reveals a forward probability measure that absorbs long-term growth rate risk. Our approach avoids the requirement of a complete set of state prices and can be used to identify sets of long-term forward probabilities that are consistent with asset pricing model restrictions and within a small deviation of rational expectations.

7   Conclusions

In this paper, we developed new methods designed to extract information on investor beliefs from data on asset prices and investor surveys. Our approach presumes an econometric model of investors or enterprises that could be misspecified under rational expectations. We illustrated how limiting the statistical discrepancy between investor beliefs and rational expectations implies bounds on investors' expectations. Formally, we represented this relationship through a nonlinear expectation function and derived its dual representation.

Going forward, we see two types of applications of our methods. Deducing market expectations about the future from forward-looking asset prices is a common practice in both the public and private sectors. But this is typically done either informally or by targeting so-called risk neutral probabilities that confound beliefs and risk preferences. Our method provides a formal way to compute and represent information on investor beliefs constrained by a model of risk aversion along with a measure of statistical divergence.

Alternatively, we could use our approach to provide diagnostics for model misspecification under rational expectations. The bounds we deduce will help assess alternative models of subjective beliefs or ambiguity aversion. Implied belief bounds for small or moderate restrictions on the statistical divergence can give suggestive results for model-builders as to how to repair potentially misspecified models. By comparing models of subjective beliefs or ambiguity aversion supported by belief distortions to the implied bounds, applied researchers could assess whether such departures from rational expectations could be easily discerned from limited data.

Future applications of our methodology could incorporate information from survey data on investor beliefs. One approach would be to include survey data directly as additional moment conditions when constructing expectation bounds. Another approach would be to compare survey-implied expectations to expectation bounds on the corresponding variables formed without using the survey data as information. The latter approach would provide a check on how plausible the survey data are as a representation of investor beliefs used in decision-making.

While we focused on constraining probabilities based on intertemporal measures of divergence, these bounds could be used formally in the design of economic policy. For instance, Woodford (2010) and Hansen and Sargent (2012) pose dynamic policy problems in which a government

policymaker fails to have precise knowledge of the beliefs of private sector when designing a prudent course of action. While these papers explore what impact this imprecision could have on policy, our work looks more systematically at how to extract credible information about the beliefs.

Acknowledgement

References

Adam, Klaus, Albert Marcet, and Juan Pablo Nicolini. 2016. Stock Market Volatility and Learning. *The Journal of Finance* 71 (1):33–82.

Alvarez, Fernando and Urban J. Jermann. 2005. Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth. *Econometrica* 73 (6):1977–2016.

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.

Antoine, Bertille, Kevin Proulx, and Eric Renault. 2018. Pseudo-True SDFs in Conditional Asset Pricing Models*. *Journal of Financial Econometrics* online.

Avellaneda, Marco. 1998. Minimum-Relative-Entropy Calibration of Asset-Pricing Models. *International Journal of Theoretical and Applied Finance* 01 (04):447–472.

Bakshi, Gurdip and Fousseni Chabi-Yo. 2012. Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors. *Journal of Financial Economics* 105 (1):191–208.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. X-CAPM: An Extrapolative Capital Asset Pricing Model. *Journal of Financial Economics* 115 (1):1–24.

Bhandari, Anmol, Jaroslav Borovicka, and Paul Ho. 2019. Survey Data and Subjective Beliefs in Business Cycle Models. Tech. rep., Federal Reserve Bank of Richmond Working Papers.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. Diagnostic Expectations and Stock Returns. *The Journal of Finance* 74 (6):2839–2874.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. Over-Reaction in Macroeconomic Expectations. *American Economic Review* forthcoming.

Borovička, Jaroslav, Lars Peter Hansen, and Jose A. Scheinkman. 2016. Misspecified Recovery. *Journal of Finance* 71 (6):2493–2544.

Bradley, Richard. 2007. *Introduction to Strong Mixing Conditions, Volume I.* Heber City, Utah: Kendrick Press.

Campbell, John. 1993. Intertemporal Asset Pricing Without Consumption Data. *American economic review* 83 (3):487—-512.

Dembo, Amir and Ofer Zeitouni. 2009. *Large Deviations Techniques and Applications*, vol. 38. Springer Science & Business Media.

Dillenberger, David, Daniel Gottlieb, and Pietro Ortoleva. 2020. Stochastic Impatience and the Separation of Time and Risk Preferences. SSRN working paper.

Donsker, Monroe D. and S.R. Srinivasa Varadhan. 1975. Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I-IV. *Communications on Pure and Applied Mathematics* 28 (1):1–47.

Duchi, John, Peter Glynn, and Hongseok Namkoong. 2018. Statistics of Robust Optimization: A Generalized Empirical Likelihood Approach.

Dupuis, Paul and Richard Ellis. 1997. *A Weak Convergence Approach to the Theory of Large Deviations.* New York: John Wiley and Sons, Inc.

Eckstein, Stephan. 2019. Extended Laplace principle for Empirical Measures of a Markov Chain. *Advances in Applied Probability* 51 (1):136–167.

Epstein, Larry G. and Stanley E. Zin. 1991. Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. *Journal of Political Economy* 99 (2):263—-286.

Epstein, Larry G., Emmanuel Farhi, and Tomasz Strzalecki. 2014. How much would you pay to resolve long-run risk? *American Economic Review* 104 (9):2680–2697.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. Natural Expectations and Macroeconomic Fluctuations. *The Journal of Economic Perspectives: a journal of the American Economic Association* 24 (4):67.

Gagliardini, Patrick and Diego Ronchetti. 2020. Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance. *Journal of Financial Econometrics* 18 (2):333–394.

Ghosh, Anisha and Guillaume Roussellet. 2020. Identifying Beliefs from Asset Prices. Tech. rep., SSRN Working paper.

Ghosh, Anisha, Christian Julliard, and Alex P. Taylor. 2017. What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. *Review of Financial Studies* 30 (2):442–504.

Hansen, Lars Peter. 1985. A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics* 30 (1-2).

———. 2014. Nobel Lecture: Uncertainty Outside and Inside Economic Models. *Journal of Political Economy* 122 (5):945 – 987.

Hansen, Lars Peter and Ravi Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *The Journal of Finance* 52 (2):557–590.

Hansen, Lars Peter and Scott F. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 55 (3):587–613.

Hansen, Lars Peter and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *The American Economic Review* 91 (2):60–66.

———. 2012. Three Types of Ambiguity. *Journal of Monetary Economics* 59 (5):422–445.

Hansen, Lars Peter and José A. Scheinkman. 2009. Long-Term Risk: An Operator Approach. *Econometrica* 77 (1):177–234.

Hansen, Lars Peter and Kenneth J. Singleton. 1996. Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors. *Journal of Business & Economic Statistics* 14 (1):53–68.

Hicks, J.R. 1939. *Value and Capital*. Oxford.

Hirshleifer, David, Jun Li, and Jianfeng Yu. 2015. Asset Pricing in Production Economies with Extrapolative Expectations. *Journal of Monetary Economics* 76:87–106.

Kazemi, Hossein B. 1992. An Intertemporal Model of Asset Prices in a Markov Economy with a Limiting Stationary Distribution. *Review of Financial Studies* 5 (1):85–104.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. London: Macmillan.

Kreps, David M. and Evan L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46 (1):185–200.

Lucas, Robert E. 1972. Expectations and the Neutrality of Money. *Journal of Economic Theory* 4 (2):103–124.

Maheu, John M. and Thomas H. McCurdy. 2000. Identifying Bull and Bear Markets in Stock Returns. *Journal of Business and Economic Statistics* 18 (1):100–112.

Manski, Charles F. 2018. Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise. *NBER Macroeconomics Annual* 32 (1):411–471.

Metzler, Lloyd A. 1941. The Nature and Stability of Inventory Cycles. *The Review of Economics and Statistics* 23 (3):113–129.

Muth, John F. 1961. Rational Expectations and the Theory of Price Movements. *Econometrica* 29 (3):315–335.

Nagel, Stefan and Kenneth J. Singleton. 2011. Estimation and Evaluation of Conditional Asset Pricing Models. *Journal of Finance* 66 (3):873–909.

Nerlove, Marc. 1958. Adaptive Expectations and Cobweb Phenomena. *The Quarterly Journal of Economics* 72 (2):227–240.

Newman, Charles M. and Bart W. Stuck. 1979. Chernoff Bounds for Discriminating Between Two Markov Processes. *Stochastics* 2:139–153.

Owen, Art B. 2001. *Empirical likelihood*. CRC Press.

Peng, Shige. 2004. Nonlinear Expectations, Nonlinear Evaluations and Risk Measures. In *Stochastic Methods in Finance: Lecture Notes in Mathematics*, edited by M. Frittelli and W. Runggaldier, 165–253. Berlin, Heidelberg: Springer Berlin Heidelberg.

Petersen, Ian R., Matthew R. James, and Paul Dupuis. 2000. Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints. *IEEE Transactions on Automatic Control* 45:398–412.

Piazzesi, Monika, Juliana Salomao, and Martin Schneider. 2020. Trend and Cycle in Bond Premia. Tech. rep., Stanford working paper.

Pigou, A.C. 1927. *Industrial Fluctuations*. Macmillan and Co. Limited, London.

Qin, Jin and Jerry Lawless. 1994. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics* 22:300–325.

Qin, Likuan and Vadim Linetsky. 2020. Long-Term Risk: A Martingale Approach. *Econometrica* 85 (1):299–312.

Ross, Steve. 2015. The Recovery Theorem. *The Journal of Finance* 70 (2):615–648.

Škulj, Damjan. 2009. Discrete Time Markov Chains with Interval Probabilities. *International Journal of Approximate Reasoning* 50 (8):1314–1329.

Stutzer, Michael. 1995. A Bayesian Approach to Diagnosis of Asset Pricing Models. *Journal of Econometrics* 68 (2):367 – 397.

———. 1996. A Simple Nonparametric Approach to Derivative Security Valuation. *The Journal of Finance* 51 (5):1633–1652.

Varadhan, S. R. 2008. Special Invited Paper: Large Deviations. *The Annals of Probability* 36 (2):397–419.

Woodford, Michael. 2010. Robustly Optimal Monetary Policy with Near-Rational Expectations. *American Economic Review* 100 (1):274–303.

Supplementary Information for
Robust Identification of Investor Beliefs

Xiaohong Chen[*]   Lars Peter Hansen[†]   Peter G. Hansen[‡]

October 27, 2020

We provide supplemental material not included in the main text in the form of appendices. Appendix A gives a simple proof that the intertemporal divergence is convex. Appendices B and C derive results about the nonlinear Perron-Frobenius which arises as the dual to the relative entropy problem. Appendix D illustrates connections between the dynamic relative entropy measure and large deviation theory. Appendix E shows how to extend the methods developed in the main text to bound ratios of expectations. Appendix F gives supplemental empirical computations. Additionally, we provide a Jupyter Notebook on `https://github.com/lphansen/Beliefs` with computational details on the implementation.

A   Convexity of the divergence

Write:

$$M_t^j = N_t^j M_{t-1}^j$$

for $j = 1, 2$. Form a convex combination:

$$\alpha M_t^1 + (1 - \alpha)M_t^2 = \left[\alpha_{t-1}N_t^1 + (1 - \alpha_{t-1})N_t^2\right]\left[\alpha M_{t-1}^1 + (1 - \alpha)M_{t-1}^2\right].$$

where

$$\alpha_{t-1} = \frac{\alpha M_{t-1}^1}{\alpha M_{t-1}^1 + (1 - \alpha)M_{t-1}^2}$$

Since ϕ is convex,

$$\mathbb{E}\left(\phi\left[\alpha_{t-1}N_t^1 + (1 - \alpha_{t-1})N_t^2\right] \mid \mathfrak{I}_{t-1}\right)$$
$$\leqslant \alpha_{t-1}\mathbb{E}\left[\phi\left(N_t^1\right) \mid \mathfrak{I}_{t-1}\right] + (1 - \alpha_{t-1})\mathbb{E}\left[\phi\left(N_t^2\right) \mid \mathfrak{I}_{t-1}\right]$$

---
[*]Yale University. Email: xiaohong.chen@yale.edu, ORCID:0000-0003-1125-675X
[†]University of Chicago. Email: lhansen@uchicago.edu, ORCID:0000-0003-0035-8558
[‡]MIT Sloan. Email: pghansen@mit.edu, ORCID:0000-0001-6351-0540

Multiplying by $\alpha M_{t-1}^1 + (1-\alpha)M_{t-1}^2$ gives:

$$\mathbb{E}\left(\phi\left[\alpha_{t-1}N_t^1 + (1-\alpha_{t-1})N_t^2\right] \mid \mathfrak{I}_{t-1}\right)\left[\alpha M_{t-1}^1 + (1-\alpha)M_{t-1}^2\right]$$
$$\leqslant \alpha\mathbb{E}\left[\phi\left(N_t^1\right) \mid \mathfrak{I}_{t-1}\right]M_{t-1}^1 + (1-\alpha)\mathbb{E}\left[\phi\left(N_t^2\right) \mid \mathfrak{I}_{t-1}\right]M_{t-1}^2$$

By taking expectations of time series averages, it follows that the intertemporal divergence is also convex:

$$\mathcal{R}\left[\alpha N_1^1 + (1-\alpha)N_1^2\right] \leqslant \alpha\mathcal{R}(N_1^1) + (1-\alpha)\mathcal{R}(N_1^1).$$

B   Perron-Frobenius problem

For an arbitrary λ_0

$$\epsilon e_0 = \mathbb{E}\left(\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0 \cdot f(X_1)\right]e_1 \mid \mathfrak{I}_0\right)$$

is recognizable as a Perron-Frobenius problem with eigenfunction e_0 and eigenvalue ϵ. The eigenfunction e_0 is in fact a random variable that is measurable with respect to $\mathfrak{I}_0$ and only well-defined up to a positive scale factor. Notice that

$$N_1 = \left(\frac{1}{\epsilon}\right)\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0 \cdot f(X_1)\right]\left(\frac{e_1}{e_0}\right)$$

is positive and has conditional expectation equal to one.

While this construction leads to a change in the one-period conditional expectation, this new probability measure does not necessarily satisfy the conditional moment restriction. The first-order conditions for minimizing with respect to λ_0, however, imply that

$$N_1^* = \left(\frac{1}{\epsilon^*}\right)\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0^* \cdot f(X_1)\right]\left(\frac{e_1^*}{e_0^*}\right)$$

satisfies:

$$\mathbb{E}\left(N_1^* f(X_1) \mid \mathfrak{I}_0\right) = 0$$

Without imposing additional regularity conditions, there may be multiple solutions to Perron-Frobenius problems. We now show that there is at most one that is pertinent to our analysis.

Lemma B.1. *When there are multiple positive eigenvalue solutions for a given λ_0, at most one of them induces a probability measure that is stochastically stable.*

Proof. By way of contradiction, we consider two possible solutions $(\tilde{\epsilon}, \tilde{e}_0)$ and $(\hat{\epsilon}, \hat{e}_0)$ where $\tilde{e}_0/\hat{e}_0$ is not constant and $\tilde{\epsilon} \geqslant \hat{\epsilon}$. Construct the corresponding $\widetilde{N}_1$ and $\widetilde{M}_T$ and let $\widetilde{\mathbb{Q}}$ be a measure-

preserving probability consistent with $\widetilde{N}_1$ and is stochastically stable. Since,

$$\widetilde{N}_1 = \left(\frac{1}{\tilde{\epsilon}}\right) \exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0 \cdot f(X_1)\right]\left(\frac{\tilde{e}_1}{\tilde{e}_0}\right),$$

it follows that

$$\mathbb{E}\left[\widetilde{N}_1\left(\frac{\hat{e}_1}{\tilde{e}_1}\right)|\mathfrak{I}_0\right] = \left(\frac{\hat{\epsilon}}{\tilde{\epsilon}}\right)\left(\frac{\hat{e}_0}{\tilde{e}_0}\right)$$

Consider two cases.

First suppose that $\tilde{\epsilon} = \hat{\epsilon}$. Then

$$\mathbb{E}\left[\widetilde{N}_1\left(\frac{\hat{e}_1}{\tilde{e}_1}\right)|\mathfrak{I}_0\right] = \frac{\hat{e}_0}{\tilde{e}_0}$$

implying that $\frac{\hat{e}_0}{\tilde{e}_0}$ is perfectly forecastable. Iterating on this relation implies that

$$\mathbb{E}\left[\widetilde{M}_T\left(\frac{\hat{e}_T}{\tilde{e}_T}\right)|\mathfrak{I}_0\right] = \frac{\hat{e}_0}{\tilde{e}_0}$$

This contradicts the presumption that $\widetilde{Q}$ is stochastically stable since the left-hand side does not converge to the corresponding unconditional expectation.

Next suppose that $\tilde{\epsilon} > \hat{\epsilon}$. Then

$$\mathbb{E}\left[\widetilde{N}_1\left(\frac{\hat{e}_1}{\tilde{e}_1}\right)|\mathfrak{I}_0\right] = \left(\frac{\hat{\epsilon}}{\tilde{\epsilon}}\right)\left(\frac{\hat{e}_0}{\tilde{e}_0}\right)$$

Since $\hat{\epsilon}/\tilde{\epsilon} < 1$, by iterating this conditional expectation operator it follows that

$$\mathbb{E}\left[\widetilde{M}_T\left(\frac{\hat{e}_T}{\tilde{e}_T}\right)|\mathfrak{I}_0\right] \to 0$$

which is contraction since $\hat{e}/\tilde{e}$ is strictly positive and cannot have a zero expectation. $\qquad\square$

C Problem solution

We solve the dual problem and verify that this satisfies the constraints for the primal problem. A comprehensive treatment of existence is beyond the scope of this paper. We will, however, provide two verification results, one for the dual and one for the primal problem.

Recall the functional equation for the dual problem:

$$\epsilon = \min_{\lambda_0} \mathbb{E}\left(\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0 \cdot f(X_1)\right]\left(\frac{e_1}{e_0}\right) | \mathfrak{I}_0\right).$$

Lemma C.1. *Let λ_0^* solve the dual problem for the eigenvalue-eigenfunction pair (ϵ^*, e_0^*). More-over, suppose that the probability measure consistent with*

$$N_1^* = \exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0^* \cdot f(X_1)\right]\left(\frac{e_1^*}{\epsilon^* e_0^*}\right)$$

and

$$M_T^* = \prod_{t=1}^{T} N_t^*$$

induces stochastic stability. Also let $\hat{\lambda}$ denote another choice of λ. Then

$$1 \leqslant \liminf_{T \to \infty} (\epsilon^*)^{-T} \, \mathbb{E}\left(\exp\left[\sum_{t=1}^{T} -\frac{1}{\xi}g(X_t) + \hat{\lambda}_{t-1} \cdot f(X_t)\right]\left(\frac{e_T^*}{e_0^*}\right) \Big| \Im_0\right)$$

Proof. Note that for all $T \geqslant 1$,

$$M_T^* \exp\left[\sum_{t=1}^{T}\left(\hat{\lambda}_{t-1} - \lambda_{t-1}^*\right) \cdot f(X_t)\right] = (\epsilon^*)^{-T} \exp\left[\sum_{t=1}^{T} -\frac{1}{\xi}g(X_t) + \hat{\lambda}_{t-1} \cdot f(X_t)\right]\left(\frac{e_T^*}{e_0^*}\right)$$

Since λ_0^* solves the dual problem,

$$
\begin{aligned}
1 &\leqslant (\epsilon^*)^{-1} \, \mathbb{E}\left(\exp\left[-\frac{1}{\xi}g(X_1) + \hat{\lambda}_0 \cdot f(X_1)\right]\left(\frac{e_1^*}{e_0^*}\right) \Big| \Im_0\right) \\
&= \mathbb{E}\left(N_1^* \exp\left[\left(\hat{\lambda}_0 - \lambda_0^*\right) \cdot f(X_1)\right] \Big| \Im_0\right)
\end{aligned}
\tag{1}
$$

Iterating on inequality (1) for T time periods gives

$$1 \leqslant \mathbb{E}\left(N_T^* \exp\left[\sum_{t=1}^{T}\left(\hat{\lambda}_{t-1} - \lambda_{t-1}^*\right) \cdot f(X_t)\right] \Big| \Im_0\right).$$

Thus

$$1 \leqslant \liminf_{T \to \infty} \mathbb{E}\left(N_T^* \exp\left[\sum_{t=1}^{T}\left(\hat{\lambda}_{t-1} - \lambda_{t-1}^*\right) \cdot f(X_t)\right] \Big| \Im_0\right)$$

with probability one. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Remark C.2. *Suppose that λ_0^* is essentially unique. Then* (1) *is satisfied with a strict inequality holding with positive probability. Under geometric ergodicity of process induced by the * probability, the right-hand-side converges to limit that is strictly greater than one.*

Next consider the primal problem.

$$\mu = \min_{N_1 \in \mathcal{N}} \mathbb{E}\left(N_1\left[g(X_1) + \xi \log N_1 + v_1\right] | \mathfrak{I}_0\right) - v_0$$

subject to the constraint:

$$\mathbb{E}\left[N_1 f(X_1) \mid \mathfrak{I}_0\right] = 0$$

To use the dual problem to construct a solution, we must verify that the first-order conditions for λ_0 are satisfied. This in turn implies that the candidate solution from the dual problem satisfies the constraint.

Lemma C.3. *Let* $(\mu^*, v_0^*, N_1^*, \lambda_0^*, \nu_0^*)$ *be constructed by solving the dual problem where* N_1^* *satisfies the constraint and* ν_0^* *is the multiplier on the constraint* $\mathbb{E}\left(N_1^* \mid \mathfrak{I}_0\right) = 1$. *Suppose that* v_0^* *is bounded. Let* $\widehat{N}_1$ *be some other change of probability measure that induces stochastic stability. Then*

$$\mu^* \leqslant \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1 + v_1^*\right] \mid \mathfrak{I}_0\right) - v_0^*.$$

Proof. From the dual problem:

$$\mu^* \leqslant \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1 + \lambda_0^* \cdot f(X_1) + v_1^* + \nu_0^*\right] \mid \mathfrak{I}_0\right) - v_0^* - \nu_0^* \qquad (2)$$

Since $\widehat{N}_1$ satisfies the the conditional moment restriction and has conditional expectation one, the Lagrange multiplier contributions drop out of:

$$\mu^* \leqslant \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1 + v_1^*\right] \mid \mathfrak{I}_0\right) - v_0^*. \qquad (3)$$

Let

$$\widehat{M}_T = \prod_{t=1}^{T} \widehat{N}_t.$$

Iterating on relation (3) gives:

$$T\mu^* \leqslant \mathbb{E}\left(\widehat{M}_T\left[\sum_{t=1}^{T} g(X_t) + \xi \log \widehat{N}_t\right] + \widehat{M}_T v_T^* \mid \mathfrak{I}_0\right) - v_0^*.$$

Dividing by T and taking limits implies that

$$\mu^* \leqslant \lim_{T \to \infty} \mathbb{E}\left(\widehat{M}_T \frac{1}{T}\left[\sum_{t=1}^{T} g(X_t) + \xi \log \widehat{N}_t\right] \mid \mathfrak{I}_0\right) + \lim_{T \to \infty} \frac{1}{T}\left[\mathbb{E}\left(\widehat{M}_T v_T^* \mid \mathfrak{I}_0\right) - v_0^*\right]$$

$$= \int \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1\right] \mid \mathfrak{I}_0\right) d\widehat{Q}_0$$

5

which follows since v_0^* is bounded and the $\hat{\cdot}$ probability induces stochastic stability. $\qquad\square$

Remark C.4. *For the Markov specification, v_0^* is a time-invariant function of only Z_0. For a bounded support set for Z_0, bounding v_0^* would seem to be widely (but not always) applicable. On the other hand, we suspect that there are weaker restrictions that would be of particular interest when the support set of Z_0 is not bounded.*

Remark C.5. *Suppose that equation (2) is satisfied with strictly positive probability. Then we may write*

$$\mu^* \leqslant \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1 + v_1^*\right] \mid \mathfrak{I}_0\right) - v_0^* - b_0$$

where the random variable $b_0 \geqslant 0$ is strictly positive with positive probability. Provided that b_0 remains strictly positive with positive probability under the $\hat{\cdot}$ probability measure,

$$\mu^* < \int \mathbb{E}\left(\widehat{N}_1\left[g(X_1) + \xi \log \widehat{N}_1\right] \mid \mathfrak{I}_0\right) d\widehat{Q}_0.$$

For the Markov specification b_0 can be written as a function of Z_0 only, so any change in measure that preserves the support of Z_0 will result in a strict inequality. This support restriction will be satisfied provided that the distorted Markov process is irreducible.

D   Statistical discrimination

In what follows, we briefly consider large deviation theory applied to empirical averages constructed in Markov settings. Consider the joint distribution of (X_t, Z_t, Z_{t-1}) which is replicated over time in accordance with a stationary and ergodic Markov process. Under the rational expectations $\mathbb{E}\left[f(X_t) \mid Z_{t-1}\right] = 0$, but without rational expectations this restriction could be violated. To assess the empirical plausibility of this restriction, select a $\lambda(Z_{t-1})$ and note that

$$\mathbb{E}\left[\lambda(Z_{t-1}) \cdot f(X_t)\right] = 0.$$

We could form a test of this restriction by checking if

$$\frac{1}{T}\sum_{t=1}^{T} \lambda(Z_{t-1}) \cdot f(X_t) \leqslant -\mathsf{c} < 0. \tag{4}$$

Of course, other tests are also possible including ones that look across a family of $\lambda(Z_{t-1})$'s and other empirical averages that include $g(X_t)$.

For a finite sample, the event (4) has positive probability under P. But the probability of this event will decline as the sample size becomes arbitrarily large. In other words, it will be increasingly rare that the expectation implied by the empirical distribution will be less than $-\mathsf{c}$.

Large deviation theory informs us about the limit

$$\frac{1}{T} \log \mathsf{P} \left\{ \sum_{t=1}^{T} \lambda(Z_{t-1}) \cdot f(X_t) < -\mathsf{c} \right\},$$

telling us how quickly these probabilities decay to zero.

The initial version of this is the type of large deviation approximation applied to empirical distributions is due to Sanov (1957) for iid sequences. It has been extended to Markov processes as discussed by Dupuis and Ellis (1997) and Varadhan (2008). Under some additional regularity conditions, remarkably, the decay can be made to be remarkably close to the minimum relative entropy over the set of possible probability measures relative to P used for representing the evolution of the (X_t, Z_t). In terms of this literature, relative entropy serves as what is called a "rate function".

We now turn to our dual formulation. When ξ is sufficiently large, the contribution of g effectively drops out of our analysis. Dropping g, gives the minimal entropy bound needed to satisfy the conditional moment restrictions. For the moment, fix λ. Large deviation theory studies the limit

$$\frac{1}{T} \log \mathsf{P} \left\{ \sum_{t=1}^{T} \lambda(Z_{t-1}) \cdot f(X_t) \geqslant 0 \mid Z_0 = z \right\}$$

Following Donsker and Varadhan, we compute the decay rate in this by solving

$$\epsilon e(z) = \min_{\mathsf{r} > 0} \mathbb{E} \left(\exp \left[\mathsf{r} \lambda(Z_0) \cdot f(X_1) \right] e(Z_1) \mid Z_0 = z \right)$$

The decay rate bound is $-\log \tilde{\epsilon}$ where $(\tilde{\epsilon}, \tilde{e})$ solve this problem. Instead of minimizing over this scalar random variable, we have multiple conditional moment conditions, and this leads us to minimize by choice of the vector function λ of the Markov realization z as a convenient way of enforcing the conditional moment restriction. The $-\log \epsilon^*$ that solves this functional equation

$$\epsilon e(z) = \min_{\lambda} \mathbb{E} \left(\exp \left[\mathsf{r} \lambda(Z_0) \cdot f(X_1) \right] e_1(Z_1) \mid Z_0 = z \right)$$

is the minimal relative entropy bound for dynamic time series evolution. When the decay rate, $-\log \epsilon^*$, is small, we view the conditional moment conditions to be particularly hard to reject.

E Bounding logarithms of ratios of expectations

To compute the lower and upper bounds on the difference in the logarithm of the expectations of two functions of X_1, we extend our previous approach as follows.

- construct $g(x) = g_1(x) - \zeta g_2(x)$ where ζ is a "multiplier" that we will search over;

7

- for alternative ζ, deduce $N_1^*(\zeta)$ and $\mathsf{Q}_0^*(\zeta)$ as described in the paper;

- compute:

$$\log \int \mathbb{E}\left[N_1^*(\zeta)g_1(X_1) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0^*(\zeta) - \log \int \mathbb{E}\left[N_1^*(\zeta)g_2(X_1) \mid \mathfrak{I}_0\right] d\mathsf{Q}_0^*(\zeta)$$

  and minimize with respect to ζ;

- set $g(x) = -g_1(x) + \zeta g_2(x)$, repeat, and use the negative of the minimizer to obtain the upper bound;

Observe that the objective is not globally convex.

F   Additional empirical computations

We compare results from relative entropy and conditional quadratic in the table that follows.

conditioning	empirical	relative entropy (lower, upper)	quadratic divergence (lower, upper)
low D/P	5.12	(1.55, 2.91)	(1.53, 2.89)
mid D/P	3.54	(2.54, 2.92)	(2.54, 2.91)
high D/P	13.89	(4.55, 5.08)	(4.41, 5.01)
unconditional	7.54	(1.74, 3.08)	(1.65, 3.03)

Table 1: Expected log market return bounds computed with relative entropy and quadratic specifications of conditional divergence. The relative entropy at the minimizing solution for the quadratic divergence is .0319. This is in comparison to .0284 when we use the relative entropy divergence.

Since our empirical method looks at other information from other excess returns, we also report bounds on other risk compensations that we included in our analysis. We convert the excess returns into returns by adding the gross returns on bonds. We report these findings in the table that follows.

conditioning	market return (lower, upper) empirical	small minus big (lower, upper) empirical	high minus low (lower, upper) empirical
low D/P	(2.63, 3.33) 5.74	(0.52, 0.94) 1.42	(-0.77, -0.01) 0.40
mid D/P	(1.84, 2.65) 3.39	(0.54, 0.87) 0.58	(-0.27, -0.08) 4.78
high D/P	(2.91, 4.25) 12.41	(0.52, 1.16) 6.44	(-1.14, -0.78) 4.99
unconditional	(2.45, 3.26) 6.84	(0.52, 0.93) 2.53	(-0.70, -0.05) 3.02

Table 2: Proportional risk compensations computed as $\log \mathbb{E}R - \log \mathbb{E}R^f$ scaled to an annualized percent. See French's data library (link) for definitions of 'small minus big' and 'high minus low'.

References

Dupuis, Paul and Richard Ellis. 1997. *A Weak Convergence Approach to the Theory of Large Deviations.* New York: John Wiley and Sons, Inc.

Sanov, I. N. 1957. On the Probability of Large Deviations of Random Magnitudes. *Matiematicheskii Sbornik* 42 (84):11–44.

Varadhan, S. R. 2008. Special Invited Paper: Large Deviations. *The Annals of Probability* 36 (2):397–419.