

ROBUST IDENTIFICATION OF INVESTOR BELIEFS

By

Xiaohong Chen, Lars Peter Hansen, and Peter G. Hansen

May 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2236



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

<http://cowles.yale.edu/>

Robust Identification of Investor Beliefs*

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

This draft: May 14, 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— Asset pricing, subjective beliefs, long-term uncertainty, ambiguity aversion, Cressie-Read divergence, generalized empirical likelihood, large deviation theory

*We thank Fernando Alvarez, Stéphane Bonhomme, Winston Dou, Darrell Duffie, Bo Honoré, Bryan Kelly, Yueran Ma, Andrey Malenko, Per Mykland, Stefan Nagel, Diana Petrova, Monika Piazzesi, Eric Renault, Azeem Shaikh, Ken Singleton, Grace Tsiang, Harald Uhlig, and Xiangyu Zhang for helpful comments and suggestions. We gratefully acknowledge the research assistance of Han Xu and Zhenhuan Xie. We thank the Alfred P. Sloan Foundation Grant [G-2018-11113] for financial support. We provide a Jupyter Notebook on <https://github.com/lphansen/Beliefs> with computational details on the implementation.

[†]Yale University. Email: xiaohong.chen@yale.edu

[‡]University of Chicago. Email: lhansen@uchicago.edu

[§]MIT Sloan. Email: pghansen@mit.edu

1 Introduction

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 the macroeconomic performance are paramount to understanding asset prices.

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 large risk prices in some time period, an arguably extreme level of estimated risk aversion, or a statistical rejection of the model. As a consequence, some researchers have explored mechanisms that could account for this evidence via a different channel, namely beliefs which differ from rational expectations. In particular, this has led to a reexamination of early models of investor beliefs such as extrapolative expectations proposed by Metzler (1941) and adaptive expectations proposed by Nerlove (1958). 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 how to go about the assignment of probabilities to future events. This has been modeled and captured formally as ambiguity aversion or concerns about model misspecification.

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. Omitting the martingale component is mathematically equivalent to an incorrect assumption of

rational expectations.

This paper proposes a formal methodology for analyzing moment restriction models where the moment restrictions are presumed to hold under a distorted probability measure. With observations on a complete set of asset prices and a known stochastic discount factor, we could identify uniquely the belief distortion. 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. Rather than estimating and testing a specific model of distorted beliefs, we study families of probabilities that are restricted to be close to rational expectations in accordance to a statistical measure of divergence.

In contrast to some earlier research applied to so-called “risk neutral probabilities,” we do not view a minimal distortion computation as a way to identify beliefs. Instead, we use it as an input into characterizing families of beliefs that have similar statistical divergences. We will be led to characterize the implications in terms of bounds. 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 characterize our bounds by building what is called a “nonlinear expectation” that minimizes expectations over members of the family of probability distortions that we identify.

When the moment restrictions are indexed by unknown parameters, instead of constructing “confidence sets” as in standard statistical analyses, we build “misspecification sets” of parameters that i) require distorted beliefs to satisfy the moment restrictions, and ii) among the permissible belief distortions are the ones that satisfy a pre-specified divergence bound from the rational expectations. Thus, our results feature the estimation of sets of models with similar magnitudes of belief distortions. The “least misspecified” specification of beliefs is a special case of our analysis.

Our paper builds on popular methods of estimation in moment restriction models such as generalized method of moments (GMM) and generalized empirical likelihood (GEL). GEL estimates parameters and probabilities jointly in hopes of improving higher-order statistical efficiency over GMM estimates. Rather than improving the statistical performance of a correctly specified moment restrictions model, we entertain and in fact feature a form of model misspecification: investor beliefs diverge from rational expectations premise that the data generating process is known to investors. In addition, some of the popular GEL

divergence criteria such as the empirical likelihood and Hellinger divergence are problematic for our analysis even though they have been used extensively for other applications of statistical approximation.¹ Moreover, we extend the use of divergence measures to be applicable to a misspecified dynamic setting pertinent to macroeconomic and financial economic applications.

In summary, we view our methodology as a way to (i) extract information on investor beliefs from asset market and survey data and (ii) to provide revealing diagnostics for model builders that embrace specific formulations of belief distortions.

1.1 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, Keynes, and Hicks, 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’s extrapolative expectations, Nerlove’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 of 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). The Adam et al. (2016) reference, in particular, points to “small departures” from rational expectations. Leading contributors to the characterization and study of martingale components of stochastic discount factors are Alvarez and Jermann (2005), Hansen and Scheinkman (2009), Bakshi and Chabi-Yo (2012), Borovička et al. (2016) Some of these contributions emphasize the explicit connection to a change in probability measure.

Our choice to focus on bounds on the potential belief distortions allows us to build on the approach of Hansen and Jagannathan (1991) and its extensions to produce meaningful

¹Schennach (2007) also demonstrates a problematic aspect of empirical likelihood but with a substantially different aim than ours.

bounds on a multiplicative component of stochastic discount factors. Precursors to our exploration of asset pricing under model misspecification include Back and Brown (1993), Stutzer (1995), Hansen and Jagannathan (1997), Luttmer et al. (1995), Almeida and Garcia (2012), Hansen (2014) and Ghosh et al. (2017). While there are overlapping ideas and motivations in this literature, there is a substantially novel component to the methods we develop.

Generalized empirical likelihood (GEL) estimators for over-identified moment restriction models have been advocated as alternatives to generalized method of moments (GMM) estimators. See, for example, the empirical likelihood alternative by Qin and Lawless (1994), and the relative entropy alternative by Imbens (1997) and Kitamura and Stutzer (1997). Both are special cases of a larger class of GEL methods based on discrepancies suggested by Cressie and Read (1984), Smith (1997) and Imbens et al. (1998). These GEL methods proceed by allowing for distortions of the empirical distribution that satisfy the moment conditions in sample, and selecting the distortion that minimizes some divergence measure. Econometricians and statisticians have defended GEL methods by showing that the resulting parameter estimators have higher-order asymptotic efficiency gains over common implementations of GMM estimators; see Newey and Smith (2004). But such methods pay little attention to enhancing our understanding of the often encountered empirical finding that there is substantial evidence against the moment conditions.² While there is an overlap between our analysis and these earlier statistics and econometrics literatures, we adopt a rather different perspective by openly acknowledging the potential global misspecification, using the implied belief distortions as barometers of private sector belief distortions, and recasting the analysis within a dynamic setting. As we will show, these differences matter in important ways.

Our approach to measuring misspecification differs in an essential way from the measures of Shanken (1987) and Hansen and Jagannathan (1997) motivated explicitly by asset pricing applications. Within the setting of linear factor models, Shanken (1987) quantifies the expected return errors induced by using misspecified proxies for a market return. Relatedly, Hansen and Jagannathan (1997) compute maximal pricing errors relative to mean-square norms of the asset payoffs. In the case of linear factor models, their measure collapses to the maximum of the expected return errors (the α 's) relative to the return standard

²When the moment conditions are misspecified, the existing GEL literature typically features the estimation of a pseudo-true parameter value, but not properties of the probabilities needed to repair the models.

deviations. In both cases, first and second moments are computed using historical data. Our analysis is different by allowing for investors to use probabilities distinct from those governing the data-generating process.

1.2 Outline of the Paper

Section 2 introduces the framework used in this paper posed as the moment restrictions implied by an asset pricing model. Section 3 studies the problem of bounding beliefs using statistical divergence constraints. We follow the GEL literature by using the Cressie and Read (1984) family of divergences between probability measures. We illustrate how some divergences may be problematic for identifying (global) misspecification. Using relative entropy divergence, we propose a nonlinear expectation functional for bounding investor expectations subject to model-implied and a divergence constraint. We also give dual representations that make evaluating the nonlinear expectation computationally tractable. Section 4 discusses the implied minimum divergence as a measure of model misspecification, and how to construct sets of parametric models consistent with small statistical divergence. We also explore the interaction of our nonlinear expectation functional with parameter identification. Section 5 gives a dynamic version of the divergence constraint motivated by large deviation theory and gives a recursive dual formulation to the corresponding nonlinear expectation. Section 6 presents an empirical illustration of our methodology. 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 $(\Omega, \mathfrak{G}, P)$ denote the underlying probability space and $\mathfrak{J} \subset \mathfrak{G}$ represent information available to investors. The original moment equations under rational expectations are of the form

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

where the function f captures the parameter dependence (θ) of either the payoff or the stochastic discount factor along with variables (X) 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{J} denote the investor information set. The underlying asset pricing equation is

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

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).³

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:

$$\mathbb{E}[Mf(X, \theta) | \mathfrak{J}] = 0. \tag{2}$$

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{J} defines a relative density that informs how rational expectations are altered by market beliefs. By changing M , we allow for alternative densities. Notice that we are restricting the implied probability measures to be absolutely continuous with respect to the original probability measure. That is, we restrict the market beliefs so that any event that is in the conditioning information set (measurable with respect to \mathfrak{J}), has probability measure zero under the original distribution will continue to have probability zero under this change in distribution. We will, however, allow for investors to assign probability zero to events that actually have positive probability under rational expectations.

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

³The vector of returns can be parameter dependent when the investment is in a physical asset with an unobserved return.

analyses. Equation (1) under rational expectations may not have solutions for any θ under rational expectations. That is, equation (1) may be misspecified. 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 identified in a rather spectacular way. The characterizing of the set of M 's associated with a given value of θ will be of particular interest to us.

Two classes of asset pricing models that have received considerable attention provide motivation for our analysis. One is that subjective beliefs 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 of asset pricing models include 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.

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 on beliefs B that reflect subjective expectations of \tilde{X} . This could include data survey responses or analyst forecasts. We may include this in our analysis by imposing the conditional moment condition:

$$\mathbb{E} \left(M \tilde{X} \mid \mathcal{J} \right) = B. \tag{3}$$

In words, this restriction says that B is the best forecast of \tilde{X} under the subjective belief measure. Note that we can incorporate probabilistic forecasts into our framework by letting \tilde{X} be an indicator function.⁴

⁴See 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

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 times series variation for these data to add nontrivial incremental information to the analysis.*

2.3 Risk-neutral pricing

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 (1996) and Avellaneda (1998) target the M 's with the smallest 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.

2.4 Martingale component to an SDF process

A stochastic discount factor process compounds the one-period stochastic discount factors. 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.

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

of over-reaction.

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.⁵

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 methods applied to this problem complement and extend those of Alvarez and Jermann (2005) and Bakshi and Chabi-Yo (2012).

2.5 Recursive utility

Consider a recursive utility model as in Kreps and Porteus (1978) and Epstein and Zin (1991). While much of the asset pricing literature appeals to a large risk aversion coefficient, we impose $\gamma = 1$ 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.

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. 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 in section 6.

⁵ It is also the measure that Ross-recovery algorithm as given in Ross (2015) produces.

Our subsequent analysis seeks to measure divergence bounds on the family of M 's that satisfy the model-implied moment conditions. We show pitfalls with some of the measures used previously, and we add methodological rigor to the associated empirical investigations.

3 Bounding beliefs

For any parameter vector θ in equation (2), there are typically many specifications of beliefs M that will satisfy the model implied moment conditions. Rather than imposing ad hoc assumptions to resolve this identification failure, we will characterize the multiplicity by using bounds on statistical divergence. A statistical divergence quantifies how close two probability measures are. In our analysis, one of these probability measures governs the data evolution while the other governs the investment decisions or the equilibrium pricing relations. We define a range of allowable probability measures, and we consider a family of divergences commonly used in the statistics literature. We then study which of these divergences are most revealing for assessing misspecification in asset pricing models. Proofs and supporting analyses for this section are given in appendix A.

For the moment, fix θ and write $f(X)$. Initially we will also abstract from the role of conditioning information, but the expectations can be interpreted as being conditioned on sigma algebra \mathfrak{J} . Later we will investigate the role of conditioning information explicitly. Introduce a convex function ϕ defined on \mathbb{R}^+ for which $\phi(1) = 0$. As a scale normalization we will assume that $\phi''(1) = 1$. The corresponding divergence of a belief M from the underlying data generation is defined by $\mathbb{E}[\phi(M)]$. By Jensen's inequality, we know that

$$\mathbb{E}[\phi(M)] \geq \phi(1) = 0$$

since $\mathbb{E}[M] = 1$. The family of divergences $\mathbb{E}[\phi(\cdot)]$ are known as f -divergences. Special cases include:

- (i) $\phi(m) = -\log m$ (negative log likelihood)
- (ii) $\phi(m) = 4(1 - \sqrt{m})$ (Hellinger distance)
- (iii) $\phi(m) = m \log m$ (relative entropy)
- (iv) $\phi(m) = \frac{1}{2}(m^2 - m)$ (Euclidean divergence).

These four cases are widely used in the GEL literature, and are nested in the family of f -divergences introduced by Cressie and Read (1984) defined by

$$\phi(m) = \begin{cases} \frac{1}{\eta(1+\eta)} [(m)^{1+\eta} - 1] & \eta < 0 \\ \frac{1}{\eta(1+\eta)} [(m)^{1+\eta} - m] & \eta \geq 0 \end{cases} \quad (4)$$

For $\eta = -1, 0$, we can apply L'Hôpital's rule to obtain cases (i) and (iii) respectively. The divergence corresponding to $\eta = -\frac{1}{2}$ is equivalent to the Hellinger distance between probability densities. Empirical likelihood methods use the $\eta = -1$ divergence. This same divergence is also featured in the analysis of Alvarez and Jermann (2005) in their characterization of the martingale component to stochastic discount factors. Two cases of particular interest to us are $\eta = 0$ and $\eta = 1$. We refer to the divergence for $\eta = 0$ as *relative entropy*. We refer to the $\eta = 1$ case as a quadratic or Euclidean divergence, which is known to have close links to GMM.

Given our interest is in sets of belief distortions, our method is distinct from those designed for estimation under correct specification. In particular, our motivation and assumptions differ substantially from the literature on GEL methods. The so-called pseudo-true parameter value that is often the centerpiece of misspecification analysis in the econometrics literature plays a tangential role in our analysis as does point identification.

3.1 Problematic divergences

For the purposes of misspecification analysis, we show that monotone decreasing divergence functions are problematic. For instance, the Cressie and Read divergences defined by (4) and used in the GEL literature are decreasing whenever $\eta < 0$. Our finding that the empirical likelihood ($\eta = -1$) and Hellinger (the case $\eta = -\frac{1}{2}$) divergences are problematic under model misspecification is noteworthy, as both have been widely used in statistics and econometrics.⁶ Our negative conclusion about monotone decreasing divergences leads us to focus on divergences for which $\eta \geq 0$ as robust measures of probability distortions.

To understand why monotone decreasing divergences are problematic, we study the corresponding population problem:

Problem 3.1.

$$\inf_{M>0} \mathbb{E}[\phi(M)]$$

⁶In particular, Hellinger distance has been used for the purpose of local robustness under misspecification.

subject to

$$\begin{aligned}\mathbb{E}[M] &= 1 \\ \mathbb{E}[Mf(X)] &= 0.\end{aligned}$$

When the constraint set is empty, we adopt the convention that the optimized objective is ∞ . We call a model misspecified if

$$\mathbb{E}[f(X)] \neq 0.$$

When f depends on an unknown parameter θ , we presume this inequality applies for all θ in a prespecified parameter space. This leads us to ask if this inequality is revealed by a strictly positive minimized objective in problem 3.1.

For a divergence to be of interest to us, the greatest lower bound on the objective should inform us as to how big of a statistical discrepancy is needed to satisfy equation (2). Therefore the infimum should be strictly positive whenever $\mathbb{E}[f(X)] \neq 0$. Conversely, notice that under correct specification, $\mathbb{E}[f(X)] = 0$, and $M = 1$ is in constraint set of problem 3.1. By the design of a divergence measure, for $M = 1$ the minimized objective for problem 3.1 is zero.

Theorem 3.2. *Assume that $\phi(m)$ is decreasing in m , $\mathbb{E}[f(X)] \neq 0$, $f(X)$ is absolutely continuous w.r.t. the Lebesgue measure on \mathbb{R}^d , and there exists a convex cone $C \subset \mathbb{R}^d$ such that $f(X)$ has strictly positive density on C and $-\mathbb{E}[f(X)] \in \text{int}(C)$. Then for any $\kappa > 0$ there exists a belief distortion M such that i) $M > 0$ on $\text{supp}[f(X)]$; ii) $\mathbb{E}[M] = 1$; iii) $\mathbb{E}[Mf(X)] = 0$; iv) $\mathbb{E}[\phi(M)] < \kappa$.*

Theorem 3.2 shows dramatically that when the vector $f(X)$ has unbounded support, problem 3.1 can become degenerate. The infimized divergence can be equal to zero even though $\mathbb{E}[f(X)] \neq 0$ so the model is misspecified. In this case the infimum is not attained by any particular M , but can be approximated by sequences that assign small probability to extreme realizations of $f(X)$.⁷ We view the assumption of unbounded support as empirically relevant, since moment conditions coming from asset pricing typically have terms that

⁷An explicit construction of such sequences is given in appendix A. Heuristically, we perturb the original distribution of $f(X)$ by shifting a very small amount of probability mass into an extreme tail so that the moment condition $\mathbb{E}[Mf(X)] = 0$ is satisfied. These perturbed distributions will converge weakly to the original distribution, and the divergence will approach zero.

are multiplicative in the returns. Note that gross returns have no a priori upper bound, and excess returns have no a priori upper or lower bounds.

The condition in Theorem 3.2 that $\phi(m)$ is decreasing in m is crucial to the degeneracy. As we noted, this condition is satisfied for the Cressie-Read family whenever $\eta < 0$.

To further understand the degeneracy for $\eta < 0$ it is helpful to consider the associated dual problem to problem 3.1. The dual problem is typically easier to solve than the primal problem, and is often the starting point for generalized empirical likelihood estimation. Consider:

$$\sup_{\lambda, \nu} \inf_{M > 0} \mathbb{E} [\phi(M) + M\lambda \cdot f(X) + \nu(M - 1)] \quad (5)$$

where λ and ν are Lagrange multipliers. Minimizing over M leads us to the dual problem:

Problem 3.3.

$$\begin{aligned} \sup_{\lambda, \nu} -\frac{1}{1+\eta} \mathbb{E} \left[(-\eta [\lambda \cdot f(X) + \nu])^{\frac{1+\eta}{\eta}} \right] - \frac{1}{\eta(1+\eta)} - \nu & \text{ if } \eta \in (-1, 0) \\ \sup_{\lambda, \nu} \mathbb{E} (\log[\lambda \cdot f(X) + \nu]) + 1 - \nu & \text{ if } \eta = -1 \end{aligned}$$

provided that $\lambda \cdot f(X) + \nu \geq 0$.

The optimized objective from problem 3.3 is necessarily less than or equal to that of the original primal problem 3.1. When the solution to the dual problem:

$$M^* = (-\eta [\lambda^* \cdot f(X) + \nu^*])^{\frac{1}{\eta}}$$

is feasible for the primal problem, then the two optimized objectives will coincide. The support restriction on $\lambda \cdot f(X) + \nu$ can be problematic under misspecification, sometimes leading to a degenerate solution of the form $\lambda^* = 0$ and $\nu^* = 1$ with the implied M^* not being feasible.

Remark 3.4. *Previously Schennach (2007) demonstrated problematic aspects of empirical likelihood estimators under misspecification. She showed that the pseudo-true parameter estimator obtained by solving the empirical likelihood estimator computed using the dual problem may fail to be root- T consistent under model misspecification, where T is the sample size. She also discusses when population dual problem may have a zero objective even though the model is misspecified. In relation to this, we showed that the primal problem may also fail to detect misspecification for any monotone decreasing divergence. This includes the $\eta = -1$ divergence used in empirical likelihood methods. As we emphasized previously, the*

methods we develop are not concerned with the point identification of pseudo-true parameter values.

3.2 Robust Bounds

We derive a nonlinear expectation functional that summarizes conveniently bounds on expectations in the presence of a divergence constraint on the probability distortion. We focus primarily on the case in which $\eta = 0$ (relative entropy divergence), although we discuss briefly the corresponding result for $\eta = 1$ (quadratic divergence).

Formally we study constrained optimization problems that will allow us to characterize either bounds on expectations of functions of the data or the divergence implied by pre-specified bounds. We initially pose these problems without reference to unknown parameters and conditioning information. We discuss both of these important extensions later in our investigation.

3.2.1 Minimal divergence

An important input into our calculations is the minimum divergence, which we now show how to compute. Formally, we solve:

Problem 3.5.

$$\underline{\kappa} = \inf_{M>0} \mathbb{E}[M \log M]$$

subject to:

$$\begin{aligned} \mathbb{E}[M f(X)] &= 0, \\ \mathbb{E}[M] &= 1. \end{aligned}$$

Instead of solving this problem directly, we investigate the conjugate or dual problem. By standard arguments, the maximized objective of the dual problem is less than or equal to the minimizing solution for problem 3.5. When the M implied by the dual problem satisfies the constraints for problem 3.5, the two optimized objectives coincide and this same M solves problem 3.5.

To formulate the dual problem, introduce two Lagrange multipliers (λ, ν) on the respective constraints.

$$\sup_{\lambda, \nu} \inf_{M>0} \mathbb{E}[M \log M + M\lambda \cdot f(X) + \nu(M - 1)].$$

As known from a variety of sources and reproduced in the appendix, the dual problem reduces to:

Problem 3.6.

$$\sup_{\lambda} -\log \mathbb{E}(\exp[-\lambda \cdot f(X)]).$$

The first-order conditions for this problem are $\mathbb{E}[M^* f(X)] = 0$ where M^* is constructed using

$$M^* = \frac{\exp[-\lambda^* \cdot f(X)]}{\mathbb{E}(\exp[-\lambda^* \cdot f(X)])}. \tag{6}$$

where λ^* is the maximizing choice of λ .

For this candidate M^* to be a valid solution, we restrict the probability distribution of $f(X)$. Notice that $\psi(\lambda) \equiv \mathbb{E}(\exp[-\lambda \cdot f(X)])$, when viewed as a function of $-\lambda$, is the multivariate moment-generating function for the random vector $f(X)$. We include $+\infty$ as a possible value of ψ in order that it be well defined for all λ . The negative of its logarithm is a concave function, which is the objective for the optimization problem that interests us. A unique solution to the dual problem exists under the following restrictions on this generating function.

Restriction 3.7. *The moment generating function ψ satisfies:*

(i) ψ is continuous in λ ;

(ii) $\lim_{|\lambda| \rightarrow \infty} \psi(\lambda) = +\infty$.⁸

A moment generating function is infinitely differentiable in neighborhoods in which it is finite. To satisfy condition (i) of restriction 3.7, we allow for ψ to be infinite as long as it asymptotes to $+\infty$ continuously on its domain. In particular, ψ does not have to be finite for all values of λ . Condition (ii) requires that ψ tends to infinity in all directions. Restriction 3.7 is satisfied when the support sets of the entries of $f(X)$ are not subsets of either the positive real numbers or negative real numbers. More importantly for us, restriction 3.7 allows for $f(X)$ to have unbounded support.

Theorem 3.8. *Suppose that restriction 3.7 is satisfied. Then problem 3.6 has a unique solution λ^* . Using this λ^* to form*

$$M^* = \frac{\exp[-\lambda^* \cdot f(X)]}{\mathbb{E}(\exp[-\lambda^* \cdot f(X)])},$$

⁸This condition rules out redundant moment conditions as well as $f(X)$'s which only take on nonnegative or nonpositive values with probability one.

this choice of M^* satisfies the two constraints imposed in problem 3.5. Thus the optimized objective for both problems is

$$\underline{\kappa} \doteq -\log \mathbb{E} \exp[-\lambda^* \cdot f(X)]$$

with M^* as the implied solution for M .

The minimal relative entropy $\underline{\kappa}$ will be a core ingredient in computations that interest us.

3.2.2 Bounding expectations

To construct misspecified sets of expectations, we use $\kappa > \underline{\kappa}$ to bound the divergence of belief misspecification. This structure will allow us to explore belief distortions other than the one implied by minimal divergence. While we represent alternative probability distributions with alternative specifications of the positive random variable M with unit expectation, we find it most useful and revealing to depict bounds on the resulting expectations. Larger κ 's will lead to bigger sets of potential expectations.

Given a function g of X , we consider the following problem:

Problem 3.9.

$$\mathbb{K}(g) \doteq \min_{M \geq 0} \mathbb{E}[Mg(X)]$$

subject to the three constraints:

$$\begin{aligned} \mathbb{E}[M \log M] &\leq \kappa \\ \mathbb{E}[Mf(X)] &= 0, \\ \mathbb{E}[M] &= 1. \end{aligned}$$

As before we can solve this problem using convex duality.⁹ The function g could define a moment of an observed variable of particular interest or it could be the product of the stochastic discount factor and an observed payoff to a particular security whose price we seek to bound.

⁹There is an extensive literature studying the mathematical structure of more general versions of this problem including more general specifications of entropy. Representatives of this literature include the insightful papers Csiszar and Matus (2012) and Csiszar and Breuer (2018). We find it pedagogically simpler to study the dual problem directly rather than to verify regularity conditions in this literature.

Consider now the set \mathcal{B} of bounded Borel measurable functions g to be evaluated at alternative realizations of the random vector X . The mapping \mathbb{K} from \mathcal{B} to the real line can be thought of as a “nonlinear expectation,” as formalized in the following proposition.

Proposition 3.10. *The mapping $\mathbb{K} : \mathcal{B} \rightarrow \mathbb{R}$ has the following properties¹⁰:*

- (i) if $g_2 \geq g_1$, then $\mathbb{K}(g_2) \geq \mathbb{K}(g_1)$.
- (ii) if g constant, then $\mathbb{K}(g) = g$.
- (iii) $\mathbb{K}(rg) = r\mathbb{K}(g)$, for a scalar $r \geq 0$
- (iv) $\mathbb{K}(g_1) + \mathbb{K}(g_2) \leq \mathbb{K}(g_1 + g_2)$

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 M 's that solve this problem can differ depending on g .

Remark 3.11. *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 .

Next we give a dual representation of $\mathbb{K}(g)$ as justified in appendix A:

$$\sup_{\xi > 0} \max_{\lambda} -\xi \log \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X) + \lambda \cdot f(X) \right] \right) - \xi \kappa. \quad (7)$$

Notice that conditioned on ξ , the maximization over λ does not depend on κ because $-\xi \kappa$ is additively separable.

It is convenient to explore the supremum over λ for each $\xi > 0$. Write:

$$\widehat{\mathbb{K}}(\xi; g) \doteq \sup_{\lambda} -\xi \log \mathbb{E} \exp \left[-\frac{1}{\xi} g(X) - \lambda \cdot f(X) \right]. \quad (8)$$

¹⁰The 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”.

We deduce ξ and the resulting moment bound by solving:

$$\mathbb{K}(g) = \sup_{\xi \geq 0} \widehat{\mathbb{K}}(\xi; g) - \xi \kappa. \quad (9)$$

Remark 3.12. *For sufficiently large values of κ used to constrain relative entropy, it is possible that this constraint actually does not bind. The additional moment restrictions by themselves limit the family of probabilities, and might do so in ways that restrict the implied entropy of the probabilities. Appendix A gives sufficient conditions under which the relative entropy constraint will bind, and provides examples suggesting that the relative entropy constraint may bind in many cases of interest even for arbitrarily large choices of κ .*

3.2.3 Alternative formulation

There is a closely related problem that is sometimes more convenient to work with. We revert back to a minimum entropy formulation and augment the constraint set to include expectations of $g(X)$ subject to alternative upper bounds. We may then deduce how changing this upper bound impacts the relative entropy objective. Stated formally,

Problem 3.13.

$$\mathbb{L}(\vartheta; g) = \inf_{M > 0} \mathbb{E} [M \log M]$$

subject to:

$$\begin{aligned} \mathbb{E} [M f(X)] &= 0, \\ \mathbb{E} [M g(X)] &\leq \vartheta \\ \mathbb{E} [M] &= 1. \end{aligned}$$

Notice that $\mathbb{L}(\vartheta; g)$ increases as we decrease ϑ because values of ϑ make the constraint set more limiting. By imitating our previous logic for the minimum divergence problem subject to moment conditions, the corresponding dual problem is:

Problem 3.14.

$$\sup_{\rho \geq 0, \lambda} -\log \mathbb{E} (\exp [-\rho g(X) - \lambda \cdot f(X)]) - \vartheta \rho.$$

The variable ρ is a Lagrange multiplier on the moment restriction involving g . We may hit a relative entropy target varying ϑ .

A natural starting point is to take the tilted solution M^* from problem 3.5 and compute

$$\mathbf{u}_g = \mathbb{E}[M^*g(X)].$$

By setting $\vartheta = \mathbf{u}_g$, the solution to problem 3.14 sets $\rho = 0$ and $\lambda = \lambda^*$. This choice satisfies the first-order conditions. Lowering ϑ will imply a binding constraint:

$$\mathbb{E}[Mg(X)] - \vartheta = 0.$$

The optimized objective gives an implied relative entropy. Given the binding constraint, we may view problem 3.13 as an extended version of problem 3.5 with an additional moment restriction added. This leads us to state following analog to theorem 3.8.

Theorem 3.15. *Suppose*

i) $\vartheta < \mathbf{u}_g$;

ii) restriction 3.7 is satisfied for the random vector: $[g(X) \quad f(X)]'$.

Then problem 3.14 has a unique solution (ρ^, λ^*) for which*

$$M^* = \frac{\exp[-\rho^*g(X) - \lambda^* \cdot f(X)]}{\mathbb{E}[\exp[-\rho^*g(X) - \lambda^* \cdot f(X)]]},$$

this choice of M^ satisfies $\mathbb{E}[M^*] = 1$, $\mathbb{E}[M^*f(X)] = 0$, and $\mathbb{E}[M^*g(X)] = \vartheta$. Thus objectives for problems 3.13 and 3.14 coincide.¹¹*

The relative entropy objective for problem 3.13 increases as we decrease ϑ . For instance, by decreasing ϑ in this way we could hit the relative entropy threshold of problem 3.9. Both approaches feature the same intermediate problem in which we initially condition on ξ or ρ and optimize over λ . For computational purposes we deduce the implied expectation of $g(X)$ and relative entropy by tracing out both as functions of the scalars ξ or ρ .

3.2.4 Bounding conditional expectations

Consider an event Λ with $P(\Lambda) = \mathbb{E}[\mathbf{1}_\Lambda] > 0$ where $\mathbf{1}_\Lambda$ is the indicator function for the event A . Given a function $g(X)$ of the data X , we can extend our previous arguments to

¹¹While ρ^*, λ^*, M^* depend on the choice of ϑ , to simplify notation we leave this dependence implicit.

produce a bound on the conditional expectation. Instead of entering $\mathbb{E}[Mg(X)] \leq \vartheta$ as an additional moment condition in problem 3.13, we include

$$\mathbb{E}(M\mathbf{1}_\Lambda [g(X) - \vartheta]) \leq 0$$

in the constraint set and vary ϑ to attain an entropy target. In practice, we solve the dual problem 3.6 as a function of ϑ tracing out the family of implied relative entropies.

3.3 Quadratic Divergence

While the $\eta = 0$ divergence has many nice properties, it imposes restrictions on thinness of tails of the probability distribution of $f(X)$ that may be too severe for some applications.¹² As an alternative, we now consider the quadratic or Euclidean divergence obtained when we set $\eta = 1$. We will not repeat the analysis of alternative bounds. Since a key input is the dual to a divergence bound problem, we will characterize the resulting solution for bounds and leave the extensions to the appendix. We study the counterpart to problem 3.9.

We impose two assumptions to ensure non-degenerate bounds.

Restriction 3.16. *$f(X)$ and $g(X)$ have finite second moments.*

Restriction 3.17. *There exists an $M > 0$ such that $\mathbb{E}[M] = 1$, $\mathbb{E}[Mf(X)] = 0$ and $\frac{1}{2}\mathbb{E}[M^2 - M] \leq \kappa$.*

The problem of interest is:

Problem 3.18.

$$\mathbb{Q}(g) \doteq \inf_{M \geq 0} \mathbb{E}[Mg(X)]$$

subject to:

$$\frac{1}{2}\mathbb{E}[M^2 - M] \leq \kappa$$

$$\mathbb{E}[Mf(X)] = 0$$

$$\mathbb{E}[M] = 1.$$

¹²For instance, if we specify S as an exponential-affine model of the form $S = \exp(\psi \cdot Z + Z' \Psi W)$ where W is a conditionally Gaussian shock, then restriction 3.7 may be violated.

We allow M to be zero with positive probability for mathematical convenience. Since there exists an $M > 0$ for which $\mathbb{E}[Mf(X)] = 0$, we can form a sequence of strictly positive M 's with divergences that are arbitrarily close to bound we derive. Solving this problem for alternative bounded g 's gives us a nonlinear expectation function \mathbb{Q} satisfying the properties in Proposition 3.10.

Problem 3.19.

$$\widehat{\mathbb{Q}}(g) \doteq \sup_{\xi \geq 0, \nu, \lambda} -\frac{\xi}{2} \mathbb{E} \left[\left(\left[\frac{1}{2} - \frac{1}{\xi} [g(X) + \lambda \cdot f(X) + \nu] \right]^+ \right)^2 \right] - \xi \kappa - \nu.$$

Proposition 3.20. *Assume that restrictions 3.16 and 3.17 hold and that the supremum in problem 3.18 is attained with $\xi^* > 0$. Then $\mathbb{Q}(g) = \widehat{\mathbb{Q}}(g)$. Furthermore, the solution $(\xi^*, \nu^*, \lambda^*)$ to problem 3.19, corresponds to the belief distortion*

$$M^* = \left[\frac{1}{2} - \frac{1}{\xi^*} [g(X) + \lambda^* \cdot f(X) + \nu^*] \right]^+$$

which satisfies the constraints of problem 3.18 with equality, and attains the infimum, i.e. $\mathbb{E}[M^*g(X)] = \mathbb{Q}(g)$.

Proposition 3.20 follows from theorem 6.7 of Borwein and Lewis (1992). It characterizes the solution to problem 3.18 when the divergence constraint binds. Otherwise, we can obtain the expectation bound by solving problem 3.19 for a fixed sequence of ξ 's converging to zero where we maximize with respect to λ and ν given any ξ in this sequence.

4 Parameter misspecification regions

This section extends our earlier analysis to accommodate parameters θ that reside in a space Θ . In our previous analysis, we took the parameter θ as given. While some of the applications we mentioned have pre-specified stochastic discount factors, many applications use stochastic discount factors that depend on unknown parameters. Moreover, in some investment-based asset pricing models, there may be unknown parameters in the implied physical or intangible returns.

Since the parameter θ now enters in the function f , we write $f(X, \theta)$ explicitly. The expectation bounds we deduced in section 3 depended implicitly on θ . Moreover, when the

expectation g is meant to be a price bound on a hypothetical asset, the function g itself may depend on θ if it is constructed using the parameterized stochastic discount factor. Alternatively, g may simply select a parameter of interest that we seek to bound.

We extend the definitions of $\underline{\kappa}$ to denote the dependence of $\theta \in \Theta$. We use this notation regardless of whether $f(X, \theta)$ satisfies restriction 3.7. When for a given θ there are no $M > 0$'s in the constraint set of problem 3.9, we adopt the convention commonly used in the convex analysis literature that $\mathbb{K}(g; \theta) = \infty$ and similarly for $\mathbb{L}(\theta)$.

4.1 Divergence bound

We start by deducing a lower bound on the divergence applicable to the entire parameter space. The entropy bound $\underline{\kappa}$ now depends on θ . Incorporating unknown parameters gives us the divergence bound is

$$\inf_{\theta \in \Theta} \underline{\kappa}(\theta). \quad (10)$$

Let $\underline{\Theta}$ be the set of minimizers. Since we now minimize over $\theta \in \Theta$, we do not need for there to be a solution to the dual problem for all θ , just for a subset of such θ 's. We could interpret $\underline{\Theta}$ as the set of “pseudo-true” parameter values coming from the relative entropy procedure proposed by Imbens (1997) and Kitamura and Stutzer (1997). Note that while the existing econometrics literature typically assumes a unique “pseudo-true” parameter value, this identification is of little concern to us. The estimation and inference on the relative entropy bound (10) and the set of minimizers $\underline{\Theta}$ could be done by applying results on partially identified nonlinear programs (or constrained M-estimation problems), see, e.g., Shapiro (1991), Chernozhukov et al. (2007), Chen et al. (2018), among others.

4.2 Parameter Uncertainty and Implied Expectations

Once we entertain the possibility of model misspecification, there is no *a priori* reason to focus only on the minimal divergence. For us, the minimum divergence problem is merely a starting point as we explore implications when $\kappa > \underline{\kappa}$, and we find it revealing to characterize implications with a nonlinear expectation operator. This leads to extend our previous construction of a nonlinear expectation to accommodate parameter uncertainty by constructing:

$$\underline{\mathbb{K}}(g) = \inf_{\theta \in \Theta} \mathbb{K}(g; \theta).$$

There is a special case of this construction that warrants special comment. Suppose

that g depends only on θ and not on X . While bounding $g(\theta)$ for a given θ is trivial, the computation of $\mathbb{K}(g)$ remains interesting as the minimum divergence conditioned on an arbitrary θ may fail to be less than or equal to κ . Notice that the dual formulation of $\mathbb{K}(g; \theta)$ simplifies to be:

$$\sup_{\xi > 0} \max_{\lambda} g(\theta) - \xi \log \mathbb{E} \exp [-\lambda \cdot f(X, \theta)] - \xi \kappa.$$

Given the linearity in ξ persists even after we maximized over λ , the outer sup problem merely checks the relative entropy constraint is satisfied or not for each θ returning an objective of $+\infty$ if it is not. This is equivalent to solving:

$$\max_{\hat{\lambda}} - \log \mathbb{E} \exp [-\hat{\lambda} \cdot f(X, \theta)]$$

for each θ ascertaining if the relative entropy constraint is satisfied or not.

Proceeding in this manner amounts to restricting a scalar function of the parameter vector θ that can be rationalized by beliefs with divergence less than some threshold. To provide a more complete characterization of a divergence set, set $g(\theta) = \pm\theta_1$ where θ_1 is the first coordinate of θ . First deduce lower and upper bounds for θ_1 , say $\underline{\theta}_1$ and $\bar{\theta}_1$. For any fixed θ_1 in the interval $(\underline{\theta}_1, \bar{\theta}_1)$ produce upper and lower bounds for θ_2 in an analogous manner.

4.3 Econometric inference

While we will develop and justify formal econometric methods in subsequent research, we now discuss briefly some similarities and differences with other econometric problems that have been studied formally. Since our starting point is belief distortion from rational expectations, we pose the estimation and inference problems differently than what researchers would under rational expectations. In contrast to some recent work in econometrics and operations research. Specifically, we take a global view of misspecification rather than a local view. Thus inferential characterizations of this misspecification are of particular interest. In contrast, for example, Duchi et al. (2018) use some very similar methods to construct robust confidence intervals for targets of estimation while localizing the impact of divergence. In light of our global perspective, inferences on the Lagrange multipliers are of particular interest as these objects are important for characterizing belief distortions that attain or approximate our bounds. These multipliers inform us how we must reshape

the historical distribution to match the moment model implications.

Second, we are interested in the implications across a set of parameter values rather than targeting a so-called pseudo-true parameter value that is often done when studying misspecification. This opens the door to connections with the extensive econometric research on set identification. For instance, we conjecture that Monte Carlo or bootstrapping techniques could provide tractable and defensible approaches for inferences. For instance, Chamberlain and Imbens (2003) and Lee (2016) propose bootstrapping methods for setups related to ours, but without a particular focus on set-based inferences. Chernozhukov et al. (2007) and Chen et al. (2018) justify Monte Carlo approximations to statistical inferences that are valid for applications with set identification. Extending such techniques to account for weakly dependent data provides a promising direction for future research.

Third, blocking approaches are often advocated as way to allow for temporal dependence when applying GEL-type methods. See, for instance, Smith (1997) and Kitamura and Stutzer (1997). It is not evident, however, that these methods are directly applicable to the problems that interest us. Blocking requires special consideration in our setting because it alters the implied measure of relative entropy; but it does in ways that are potentially interpretable. In applications, however, it is typically the distorted conditional distributions or conditional moments that are of interest. As we will see in the next section, this leads us to an extension that applies recursive methods of optimization familiar from dynamic programming as an alternative way to confront the time series structure in the data.

5 Intertemporal Divergence

Asset pricing models imply conditional moment restrictions. This leads us to explore formally the impact of conditioning. In this section we propose and justify a dynamic extension that a) accommodates conditioning and b) uses the recursive structure of multi-period likelihoods.

The most direct extension of our static formulation in section 3, would simply apply the analysis in the section 3 conditioned on sigma algebra \mathfrak{J} including both the objective and the constraints. However, there are two interrelated limitations to this approach. First, the minimum divergence will depend on the conditioning information as well as the entropy constraints once we push away from the minimum. Second, the only probabilities that would be distorted are the conditional probabilities and *not* the probabilities over the conditioning information. In a dynamic environment, distorting one-period transition

probabilities also alters the probabilities of conditioning information in the next period in ways that are inconsistent with probabilities over conditioning information in the current period. This inconsistency leads us to pose a dynamic counterpart to the analysis in Section 3 that imposes consistency requirements between conditional probabilities that govern transitions and the probabilities of the future conditioning information.

In this section we construct and use the dynamic counterpart to the statistical divergence constraint. By focusing initially on the case in which $\eta = 0$, we adopt a notion of relative entropy which frequently arises in the analysis of large deviations of stochastic processes with temporal dependence.¹³ As we will show, our application of relative entropy as formulated in this section has a direct interpretation in terms of statistical discrimination for broad classes of temporally dependent processes. In particular, it can be viewed as the *rate* at which the two probability measures become statistically discernible. We also describe briefly how to extend other divergences in analogous ways.

5.1 Data generation

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.

We start with a baseline probability triple $(\Omega, \mathfrak{G}, \mathbf{P})$ and a measurable one-to-one transformation \mathbb{U} which is measure-preserving and ergodic under \mathbf{P} . We use \mathbb{U} to construct stochastic processes and filtrations.¹⁴

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:

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

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

¹³See, for instance, Donsker and Varadhan (1975) or Dupuis and Ellis (1997).

¹⁴A common specification of \mathbb{U} is the shift transformation applied to the space of infinite sequences of vectors of real numbers.

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

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

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

5.2 Alternative probabilities

Let \mathbb{Q} denote an alternative probability distribution on (Ω, \mathfrak{G}) that is measure-preserving and ergodic, and let \mathbb{Q}_t be the restriction of \mathbb{Q} to \mathfrak{I}_t . We consider only \mathbb{Q} 's for which there exists an $N_1 \geq 0$ that is \mathfrak{I}_1 measurable and satisfies:

$$\int B_1 d\mathbb{Q}_1 = \int \mathbb{E}(N_1 B_1 | \mathfrak{I}_0) d\mathbb{Q}_0 \tag{11}$$

for all bounded \mathfrak{I}_1 measurable random variables B_1 . This N_1 necessarily satisfies $\mathbb{E}(N_1 | \mathfrak{I}_0) = 1$.

Form the product

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

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

$$\mathbb{E}(M_T B_T | \mathfrak{I}_0).$$

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 \mathbb{Q} to imply stochastic stability:

Definition 5.1. *We say that \mathbb{Q} induces **stochastic stability** if for any B_0 that is \mathfrak{I}_0 measurable and satisfies $\int |B_0| d\mathbb{Q}_0 < \infty$,¹⁵*

$$\lim_{T \rightarrow \infty} \mathbb{E}(M_T B_T | \mathfrak{I}_0) = \int B_0 d\mathbb{Q}_0.$$

¹⁵Stochastic 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 5.2. *The set \mathcal{N} contains all N_1 's for which they correspond a probability \mathbf{Q} satisfying (11) and is stochastically stable.*

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

5.3 Likelihoods

We represent the expected log-likelihood ratio as a sum of contributions for each date by using the recursive structure of a relative likelihood:

$$\mathbb{E}(M_T \log M_T | \mathfrak{J}_0) = \mathbb{E}\left(M_T \sum_{t=1}^T \log N_t | \mathfrak{J}_0\right) \geq 0.$$

Dividing by T and taking limits gives:

$$\mathcal{R}(N_1) \doteq \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}(M_T \log M_T | \mathfrak{J}_0) = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}\left(M_T \sum_{t=1}^T \log N_t | \mathfrak{J}_0\right) = \int \mathbb{E}(N_1 \log N_1 | \mathfrak{J}_0) d\mathbf{Q}_0,$$

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 \mathbf{Q}_0 , which gives rise to a restriction that we impose when computing bounds.

Remark 5.3. *The relative entropy measure $\mathcal{R}(N_1)$ is the discrete-time analog to the relative entropy measure that is used in the Donsker-Varadhan large deviation theory applied to Markov processes.¹⁶*

Remark 5.4. *Using positive random variables, M_T , to depict alternative probabilities for date T events imposes absolute continuity (conditioned on date zero information). This same absolute continuity will not be true over the infinite future. Our division by T when constructing relative entropy purposefully allows for the the altered probability to have different Law of Large Numbers limits.*

5.4 Moment bounds

In formulating the computation, 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 relative entropy penalty indexed by a parameter $\xi > 0$ and show

¹⁶See Dupuis and Ellis (1997) and Varadhan (2008) and Dembo and Zeitouni (2009) (see chapter 3).

how to solve a problem given ξ . We will then treat ξ as a Lagrange multiplier and trace out the implied relative entropies for each such ξ . In this way, ξ may be chosen to enforce a constraint. For notational simplicity we suppress the parameter dependence.

Problem 5.5.

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

subject to the constraint:

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

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

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 following objects are of interest from this problem:

- the moment bound: $\int \mathbb{E} [N_1^* g(X_1) | \mathfrak{I}_0] d\mathbf{Q}_0^*$;
- the corresponding conditional moment: $\mathbb{E} [N_1^* g(X_1) | \mathfrak{I}_0]$;
- the implied relative entropy: $\int \mathbb{E} [N_1^* \log N_1^* | \mathfrak{I}_0] d\mathbf{Q}_0^*$

where N_1^* solves problem 5.5 and \mathbf{Q}_0^* is an implied stationary distribution.

There are two features of problem 5.5 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 . Second, $\xi \mathbb{E} (N_1 \log N_1 | \mathfrak{I}_0)$ acts as a per period relative entropy penalty, but we may equivalently think of ξ as a Lagrange multiplier and subsequently maximize over ξ . We elaborate on both of these points in the discussion that follows.

We motivate the μ and v_0 in functional equation in problem 5.5 as follows. Let N_1^* be the solution and write:

$$\mu^* = \mathbb{E} (N_1^* [g(X_1) + \xi \log N_1^* + v_1^* - v_0^*] | \mathfrak{I}_0).$$

Let \mathbf{Q}_0^* be the implied stationary measure, and

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

By iterating on functional equation, we find that

$$T\mu^* = \mathbb{E} \left(M_T^* \sum_{t=1}^T [g(X_t) + \xi \log N_t^*] \mid \mathcal{I}_0 \right) + \mathbb{E} (M_T^* v_T^* \mid \mathcal{I}_0) - v_0^*.$$

Provided that $g(X_t) + \xi \log N_t^*$ has a finite unconditional expectation under \mathbb{Q} , then the same is true of v_0^* , and

$$\lim_{T \rightarrow \infty} \mathbb{E} (M_T^* v_T^* \mid \mathcal{I}_0) = \int v_0^* d\mathbb{Q}_0^*.$$

It follows that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left(M_T^* \sum_{t=1}^T [g(X_t) + \xi \log N_t^*] \mid \mathcal{I}_0 \right) = \int \mathbb{E} (N_1^* [g(X_1) + \xi \log N_1^*] \mid \mathcal{I}_0) d\mathbb{Q}_0^* = \mu^*.$$

Thus μ^* is the mean of $[g(X_t) + \xi \log N_t^*]$ under the distorted probability measure. This minimization problem aims to make μ^* as small as possible where N_1 is restricted to satisfy the conditional moment conditions.

To compute the bound μ^* requires that we take into account that the choice of N_1 has implications for future time periods. We capture this by the continuation values v_1 and v_0 represented as random variables. If we subtract this mean, we obtain a more refined result:

$$v_0^* - \int v_0^* d\mathbb{Q}_0^* = \lim_{T \rightarrow \infty} \mathbb{E} \left(M_T^* \sum_{t=1}^T [g(X_t) + \xi \log N_t^* - \mu^*] \mid \mathcal{I}_0 \right).$$

Notice that the random variable v_0^* is only determined up to a translation. We incorporate v_1 in the minimization because the choice of N_1 has implications for the beliefs about future values of the objective function that are present when we solve the expectational difference equation forward.

While we formulated problem 5.5 in terms of the parameter ξ , as we noted previously, we may interpret it as a Lagrange multiplier. Thus, we use ξ to index alternative problems that penalize the increment to relative entropy. The value μ of this objective depends on ξ leading us to write $\mu^*(\xi)$. To impose a specific relative entropy constraint κ , we solve

$$\sup_{\xi \geq 0} \mu^*(\xi) - \xi \kappa.$$

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 relative entropy associated with N_1^* . To determine the ξ sensitivity, many versions of the optimization problem could be solved in parallel using the dual approach that we next describe.

5.5 Dual problem

We pose the dual problem as a principal eigenvalue problem. This gives a revealing representation of the distorted measures that underlies the bounds and an revealing link to large deviation theory.

By imitating our earlier application of duality,

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

where λ_0 is restricted to be \mathfrak{J}_0 measurable. Let $\epsilon = \exp \left(-\frac{\mu}{\xi} \right)$ and $e_0 = \exp \left(-\frac{v_0}{\xi} \right)$. Then a equivalent statement of the dual problem is:

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

In this optimization problem λ_0 is again restricted to be a \mathfrak{J}_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, at most one of these is of interest to us.

Lemma 5.6. *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 appendix B for a proof.¹⁷

Proposition 5.7. *Problem 5.5 can be solved by finding the solution to:*

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

¹⁷Hansen and Scheinkman (2009) prove a counterpart to this result for continuous-time specifications in a Markovian environment. Qin and Linetsky (2020) extend their analysis by, among other things, relaxing the Markov assumption.

where

$$\begin{aligned}\mu &= -\xi \log \epsilon \\ v_0 &= -\xi \log e_0.\end{aligned}$$

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:

$$N_1^* = \frac{\exp\left[-\frac{1}{\xi}g(X_1) + \lambda_0^*(Z_0) \cdot f(X_1)\right] e_1^*}{\epsilon^* e_0^*}$$

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

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

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

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

The implied relative entropy is

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

Formally, the computed bound is for the unconditional expectation of $g(X_1)$, although we find the conditional expectation, $\mathbb{E}[N_1^* g(X_1) \mid \mathfrak{I}_0]$ that is a central part of the calculation, to be of interest in its own right. Alternatively, given a discrete representation of the conditioning information, we could produce a bound analogous to that in section 3.2.4.

In our statement of problem 5.5, we suppressed the parameter dependence. Provided that we can compute the solution of the dual problem quickly, we can assess parameter sensitivity along the lines we mentioned previously by perhaps solving this problem many times in parallel.

5.6 Markov specification

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

Assumption 5.8. $\{(X_t, Z_t) : t = 0, 1, \dots\}$ 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

$$e(z) = \min_{\lambda} \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X_1) + \lambda(z) \cdot f(X_1) \right] e(Z_1) \mid Z_0 = z \right) \quad (12)$$

where by an abuse of notation, we now let e and λ be functions of the Markov state Z_t with z denoting a potential realized value. 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.

5.7 Connections to large deviation theory

Our characterization is reminiscent of the results from large deviation theory for Markov processes. Large deviation theory also includes dynamic optimization in conjunction with a Laplace principle, making the two literatures closely related. As in our analysis, large deviation theory studies an undiscounted limiting problem. See, for instance, Dupuis and Ellis (1997) and Varadhan (2008) for a valuable treatise on large deviation theory.¹⁸

We investigate a more substantive link to large deviations that helps us interpret the relative entropy bounds that we input into our analysis. Suppose we use the empirical probability to detect potential departures from the baseline model. There is typically a

¹⁸In 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.

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. Large deviation theory characterizes this rate. Under some additional regularity conditions, remarkably, the decay can be made to be arbitrarily close to the minimum relative entropy bound that we compute. More generally, relative entropy provides what is called a “rate function.” Of particular interest to us, it computes the small excursions, represented probabilistically, that make decay rate as small as possible. See appendix B for an elaboration.

While we draw on insights from large deviation theory, our ultimate aim is quite different from that theory. We characterize families of probabilities that correct the misspecification induced by imposing a constraint on the divergence from rational expectations. This is in contrast to the large deviation ambition of characterizing the most likely excursions from a baseline probability measure.

5.8 Alternate divergences

So far, we have imposed the relative entropy divergence. Relative entropy limits tail behavior of the probability distributions. For this reason we consider other divergences, choices of $\eta > 0$ and, in particular, $\eta = 1$. Indeed problem 5.5 has a counterpart when $\eta > 0$, but it requires some qualification and a dynamic extension that continues to exploit the recursive structure implied by the likelihood factorization.

Specifically, the analysis extends when we use:

$$\mathcal{R} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} (M_t \mathbb{E} [\phi(N_{t+1}) | \mathfrak{I}_t] | \mathfrak{I}_0).$$

where ϕ is a convex function used as a discrepancy for the one-period transition probability. The limiting version of this measure as implied by the Law of Large Numbers for stationary, ergodic processes is:

$$\mathbb{E} [\phi(N_1) | \mathfrak{I}_0] d\mathbb{Q}.$$

With this construction, our analysis for the $\eta = 0$ case extends.

Problem 5.9. *Find a pair (μ, v) that satisfy*

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

subject to:

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

where N_1 is \mathfrak{I}_1 measurable and $v_1 = v_0 \circ \mathbb{U}$.

As before, this problem can be solved with convex duality methods.¹⁹

Remark 5.10. *Eckstein (2019) has extended the Laplace principle from Large Deviation theory include a class of intertemporal divergences that include the ones we use here. While he poses his problem with sufficient generality that constraints could be included, his formulation does not nest the restrictions of interest to us.*

6 Illustration

Consider a recursive utility model as in Epstein and Zin (1991) with risk aversion $\gamma = 1$. We impose a moderate risk aversion and instead explore belief distortions to explain the observed heterogeneity in expected returns. Let R^w denote a presumed observable return on wealth. As noted in section 2.5, under distorted beliefs represented by M ,

$$S = M(R^w)^{-1} \tag{13}$$

where S is the one-period stochastic discount factor under rational expectations.

Epstein and Zin (1991) note the additional consumption Euler equation²⁰

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

where G is the ratio of consumption growth over two adjacent time periods. 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.

¹⁹Problems 5.5 and 5.9 are closely related to the value-function recursions coming from moment-constrained variational preferences proposed by Hansen (2020).

²⁰See equations (17) and (18) of Epstein and Zin (1991) except that we allow for more general expectations.

Conversely, they are negatively related when $\rho > 1$.²¹ 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 known to be persistent, and this will be evident in our calculations.²²

We implement our approach using quarterly data from 1954-2016. We proxy for the return on wealth using the return on CRSP value-weighted index. 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 (13), 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 5 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 \bullet '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. Interestingly, it is when we condition on the low value of the dividend-price ratio

²¹This follows by taking an approximation of the logarithm of the wealth-consumption ratio around $\rho = 1$.

²²As 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.

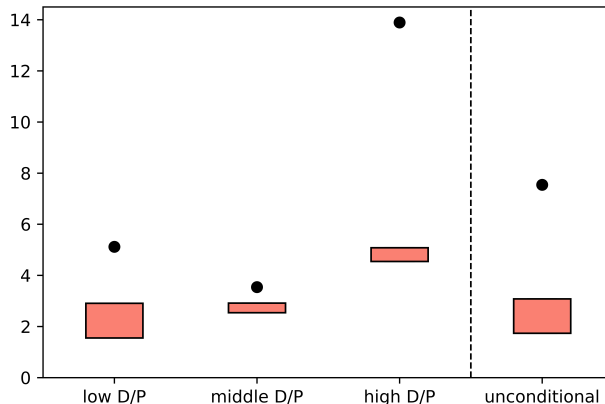


Figure 1: Expected log market return. The •’s are empirical averages and the boxes give the imputed bounds when we inflated the minimum relative entropy by 20%. The minimum relative entropy is .0284 with a half-life of 24.4 quarters.

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.

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 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 transition and depart from the empirical transition matrix in comparable ways.

There is a substantial asset pricing literature that studies time-varying risk compensation, often appealing to high values of risk aversion. Belief distortions can imitate compo-

	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	[.42 .31 .27]	[.76 .20 .04]

Table 1: Empirical and distorted transition probabilities.

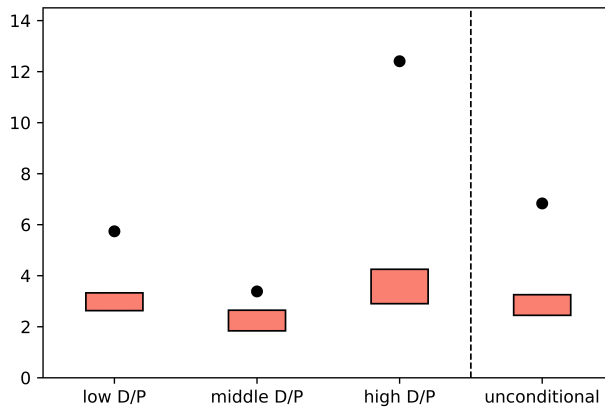


Figure 2: Proportional risk compensations computed as $\log \mathbb{E}R^w - \log \mathbb{E}R^f$ scaled to an annualized percent. The •’s are the empirical averages and the boxes give the imputed bounds when we inflated the minimum relative entropy by 20%. The minimum relative entropy is .0284 with a half-life of 24.4 quarters.

nents of this risk aversion. Thus, consider the bounds on the implied risk compensations when we restrict the risk aversion parameter to 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 5 as described in the appendix C. Not surprisingly, restricting investor risk aversion allows for belief distortions to capture the fluctuating empirical compensations for exposure to uncertainty.

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 are 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.

7 Conclusion

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 investor risk aversion, one 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. Additionally, we showed how to use the implied minimal statistical divergence as a measure of model misspecification, and discussed how to estimate sets of parametric models that can be rationalized by small statistical departures from rational expectations.

Our implementation uses empirical distributions in analogous manners as GMM and GEL methods with sparse clustering of conditioning information. Sieve methods could open the door to richer specifications of this information. For instance, our approach could also be applied in vector autoregressive settings using log-linear or even higher order approximation of structural models. Macroeconomic researchers sometimes introduce ad hoc belief distortions relative to rational expectations to repair the structured models. Alternatively, reduced-form models with hidden Markov processes could be taken as the empirical starting point.

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 our paper makes reference to prior inferential results that can be used, we believe it will be fruitful to develop these links more formally in future research. Of particular interest are inferential methods for the nonlinear expectation function implied by the dynamic relative entropy constraint using flexible conditioning information. In this regard, inferential methods for sieve-type estimation as surveyed by Chen (2007) could be adapted or extended to our setting. For instance, Christensen (2017) has applied this approach to the estimation of principal eigenvalues and eigenfunctions, and extensions could be applied to our dual formulation in a dynamic setting. In a different vein, the extended version of GMM estimation developed in Gagliardini et al. (2011) allows for applied problems in which a researcher wishes to use cross-sectional richness from asset markets in some of the observed time periods. Modifying this approach to allow for investor belief distortions would be another fruitful avenue to explore.

References

- Adam, Klaus, Albert Marcet, and Juan Pablo Nicolini. 2016. Stock Market Volatility and Learning. *The Journal of Finance* 71 (1):33–82.
- Almeida, Caio and Rene Garcia. 2012. Assessing Misspecified Asset Pricing Models with

- Empirical Likelihood Estimators. *Journal of Econometrics* 170 (2):519 – 537.
- 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.
- Avellaneda, Marco. 1998. Minimum-Relative-Entropy Calibration of Asset-Pricing Models. *International Journal of Theoretical and Applied Finance* 01 (04):447–472.
- Back, Kerry and David P. Brown. 1993. Implied Probabilities in GMM Estimators. *Econometrica* 61 (4):971–975.
- 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.
- 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 José A. Scheinkman. 2016. Misspecified Recovery. *The Journal of Finance* 71 (6):2493–2544.
- Borwein, Jonathan M. and Adrian S. Lewis. 1992. Partially Finite Convex Programming, Part II: Explicit Lattice Models. *Mathematical Programming* 57 (1-3):49–83.
- 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.
- Chamberlain, Gary and Guido W. Imbens. 2003. Nonparametric Applications of Bayesian Inference. *Journal of Business and Economic Statistics* 21 (1):12–18.

- Chen, Xiaohong. 2007. Chapter 76: Large Sample Sieve Estimation of Semi-Nonparametric Models. In *Handbook of Econometrics*, vol. 6, Part B, edited by James J. Heckman and Edward E. Leamer, 5549 – 5632. Elsevier.
- Chen, Xiaohong, Tim Christensen, and Elie Tamer. 2018. Monte Carlo Confidence Sets for Identified Sets. *Econometrica* 86 (6):1965–2018.
- Chernozhukov, V., H. Hong, and E. Tamer. 2007. Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica* 75 (5):1243–1284.
- Christensen, Timothy M. 2017. Nonparametric Stochastic Discount Factor Decomposition. *Econometrica* 85 (5):1501–1536.
- Cressie, Noel and Timothy R. C. Read. 1984. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society. Series B (Methodological)* 46 (3):440–464.
- Csiszar, I. and Thomas Breuer. 2018. Expected Value Minimization in Information Theoretic Multiple Priors Models. *IEEE Transactions on Information Theory* 64 (6):3957–3974.
- Csiszar, I. and F. Matus. 2012. Generalized Minimizers of Convex Integral Functionals, Bregman Distance, Pythagorean Identities. *Kybernetika* 48 (4):637–689.
- Dembo, Amir and Ofer Zeitouni. 2009. *Large Deviations Techniques and Applications*, vol. 38. Springer Science & Business Media.
- 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.

- Ekeland, I. and R. Témam. 1999. *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.
- 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.
- 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, Christian Gourieroux, and Eric Renault. 2011. Efficient Derivative Pricing by the Extended Method of Moments. *Econometrica* 79 (4):1181–1232.
- 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. 2014. Nobel Lecture: Uncertainty Outside and Inside Economic Models. *Journal of Political Economy* 122 (5):945 – 987.
- Hansen, Lars Peter and Ravi Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99 (2):225–62.
- . 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *The Journal of Finance* 52 (2):557–590.
- Hansen, Lars Peter and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *The American Economic Review* 91 (2):60–66.
- Hansen, Lars Peter and José A. Scheinkman. 2009. Long-Term Risk: An Operator Approach. *Econometrica* 77 (1):177–234.
- Hansen, Peter G. 2020. New Formulations of Ambiguous Volatility with an Application to Optimal Dynamic Contracting. Working paper.
- Hirshleifer, David, Jun Li, and Jianfeng Yu. 2015. Asset Pricing in Production Economies with Extrapolative Expectations. *Journal of Monetary Economics* 76:87–106.

- Imbens, Guido W. 1997. One-Step Estimators for Over-Identified Generalized Method of Moments Models. *The Review of Economic Studies* 64 (3):359–383.
- Imbens, Guido W., Richard H. Spady, and Phillip Johnson. 1998. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* 66 (2):333–357.
- 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.
- Kitamura, Yuichi and Michael Stutzer. 1997. An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica* 65 (4):861–874.
- Kreps, David M. and Evan L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46 (1):185–200.
- Lee, Seojeong. 2016. Asymptotic Refinements of a Misspecification-Robust Bootstrap for GEL estimators. *Journal of Econometrics* 192:86–104.
- Lucas, Robert E. 1972. Expectations and the Neutrality of Money. *Journal of Economic Theory* 4 (2):103–124.
- Luttmer, Erzo, Lars P. Hansen, and John Heaton. 1995. Econometric Evaluation of Asset Pricing Models. *Review of Financial Studies* 8:237–274.
- 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: Journal of the Econometric Society* 315–335.
- Nerlove, Marc. 1958. Adaptive Expectations and Cobweb Phenomena. *The Quarterly Journal of Economics* 227–240.
- Newey, Whitney K. and Richard J. Smith. 2004. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72 (1):219–255.

- 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.
- Qin, Jin and Jerry Lawless. 1994. Empirical Likelihood and General Estimating Equations. *Ann. Statist.* 22 (1):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.
- Sanov, I. N. 1957. On the Probability of Large Deviations of Random Magnitudes. *Matematicheskii Sbornik* 42 (84):11–44.
- Schennach, Susanne M. 2007. Point Estimation with Exponentially Tilted Empirical Likelihood. *Annals of Statistics* 35 (2):634–672.
- Shanken, Jay. 1987. Multivariate Proxies and Asset Pricing Relations. *Journal of Financial Economics* 18 (1):91 – 110.
- Shapiro, Alexander. 1991. Asymptotic Analysis of Stochastic Programs. *Annals of Operation Research* 30:169—186.
- Smith, Richard. 1997. Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation. *Economic Journal* 107:503–519.
- 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.

Online Appendix

A Proofs and Derivations for Section 3

A.1 Proof of Theorem 3.2

Construct a sequence $\pi_j \searrow 0$ such that $\pi_j < \frac{1}{2}$ for all j . Then choose $r_j \in \mathbb{R}^d$ such that

$$(1 - \pi_j)\mathbb{E}[f(X)] + \pi_j r_j = 0$$

i.e.

$$r_j = - \left(\frac{1 - \pi_j}{\pi_j} \right) \mathbb{E}[f(X)]$$

Let $B(r, \epsilon)$ denote an open ball with center r and radius ϵ . Since $-\mathbb{E}[f(X)] \in \text{int}(C)$ there exists an $\epsilon > 0$ such that the open ball $B(-\mathbb{E}[f(X)], \epsilon) \subset C$. Since C is a cone and $\pi_j < \frac{1}{2}$ it follows that $B(r_j, \epsilon) \subset C$. Write $v(\epsilon) = \text{vol}[B(0, \epsilon)] > 0$.²³ Now, construct a sequence of belief distortions M_j as follows:

$$M_j(x) = (1 - \pi_j) + \pi_j \frac{1}{v(\epsilon)h_0[f(x)]} \mathbf{1}\{f(x) \in B(r_j, \epsilon)\}$$

where $h_0(y)$ is the density of the random variable $Y = f(X)$ under the objective probability measure P . By construction, we have that for all $j \in \mathbb{N}$

- $M_j > 0$
- $\mathbb{E}[M_j] = 1$
- $\mathbb{E}[M_j f(X)] = 0$.

Additionally note that $M_j \geq (1 - \pi_j)$ with probability one. Since $\phi(\cdot)$ is decreasing, we have that $\phi(M_j) \leq \phi(1 - \pi_j)$ with probability one. By continuity, $\phi(1 - \pi_j) \rightarrow \phi(1) = 0$. By monotonicity of expectations we see that

$$0 \leq \mathbb{E}[\phi(M_j)] \leq \mathbb{E}[\phi((1 - \pi_j))] = \phi(1 - \pi_j) \rightarrow 0.$$

The statement follows immediately. □

²³Here we use the definition $\text{vol}(S) = \int \mathbf{1}(y \in S) dy$.

A.2 Derivation of Problem 3.3

By standard duality arguments, the dual formulation of problem 3.1 is the saddlepoint equation

$$\sup_{\lambda, \nu} \inf_{M > 0} \mathbb{E} [\phi(M) + M\lambda \cdot f(X) + \nu(M - 1)] \quad (14)$$

where λ and ν are Lagrange multipliers.

The objective function is separable over the realized values of M , and this leads us to minimize:

$$\phi(M) + M\lambda \cdot f(X) + \nu(M - 1)$$

The first-order condition for optimizing over M is:

$$\frac{1}{\eta} M^\eta + \lambda \cdot f(X) + \nu = 0.$$

Thus the minimizing M is

$$M = (-\eta[\lambda \cdot f(X) + \nu])^{\frac{1}{\eta}}$$

Substituting the minimizing M back into (5) leads us to

$$-\left(\frac{1}{1+\eta}\right) M^{\eta+1} - \frac{1}{\eta(1+\eta)} - \nu = -\left(\frac{1}{1+\eta}\right) (-\eta[\lambda \cdot f(X) + \nu])^{\frac{\eta+1}{\eta}} - \frac{1}{\eta(1+\eta)} - \nu$$

as in dual problem 3.3.

A.3 Proof of Theorem 3.8

The negative of a log moment generating function is strictly concave. Conditions (i) and (ii) guarantee that the function ψ is continuous and coercive. It follows from (Ekeland and Témam, 1999, Proposition 1.2, Ch. II.1, p.35) that the supremum in Problem 3.6 is attained uniquely at vector we denote λ^* . Since ψ is differentiable, λ^* is determined uniquely by solving the first-order conditions. Moreover, from known results about moment generating functions we may differentiate inside the expectation to conclude that the first-order conditions with respect to λ imply

$$\mathbb{E} \left[\frac{\exp(\lambda^* \cdot f(X))}{\mathbb{E}[\exp(\lambda^* \cdot f(X))]} f(X) \right] = \mathbb{E}[M^* f(X)] = 0.$$

This can be seen directly via the dominated convergence theorem. Thus M^* is feasible for Problem 3.5.

To verify that M^* solves Problem 3.5, note that for any other $M \geq 0$ with $\mathbb{E}[M] = 1$,

$$\mathbb{E}[M(\log M - \log M^*)] \geq 0,$$

and thus

$$\mathbb{E}[M \log M] \geq \mathbb{E}[M \log M^*].$$

The first expression is nonnegative because it is the entropy of M relative to M^* .²⁴ Compute

$$\mathbb{E}[M \log M^*] = \mathbb{E}[M(\lambda^* \cdot f(X))] - \log \mathbb{E}[\exp(\lambda^* \cdot f(X))].$$

Thus if $\mathbb{E}[Mf(X)] = 0$,

$$\mathbb{E}[M \log M^*] = -\log \mathbb{E}[\exp(\lambda^* \cdot f(X))].$$

We conclude that

$$\inf_{\mathbf{B}} \mathbb{E}[M \log M] \geq -\log \mathbb{E}[\exp(\lambda^* \cdot f(X))]$$

where $\mathbf{B} = \{M \in L^1(\Omega, \mathfrak{G}, P) : \mathbb{E}[M] = 1, \mathbb{E}[Mf(X)] = 0\}$. Furthermore, the right-hand side is attained by setting $M = M^*$ and that other $M \in \mathbf{B}$ that attains the infimum is equal to M^* with probability one. \square

A.4 Derivation of equation (7)

By standard duality arguments, the dual formulation of problem 3.9 is the saddlepoint equation

$$\sup_{\xi > 0, \lambda, \nu} \inf_{M \geq 0} \mathbb{E}[Mg(X) + \xi(M \log M - \kappa) + \lambda \cdot Mf(X) + \nu(M - 1)] \quad (15)$$

where ξ, λ and ν are Lagrange multipliers. Since the objective function is separable in M , we minimize

$$Mg(X) + \xi(M \log M - \kappa) + \lambda \cdot Mf(X) + \nu(M - 1)$$

²⁴Formally $\mathbb{E}[M(\log M - \log M^*)] = \mathbb{E}[M^* \phi(M/M^*)]$ with $\phi(x) = x \log x$, so the expectation is non-negative by Jensen's inequality.

with respect to M . The first-order condition is

$$g(X) + \xi + \xi \log M + \lambda \cdot f(X) + \nu = 0.$$

Thus,

$$M = \frac{\exp\left(-\frac{1}{\xi} [g(X) + \lambda \cdot f(X)]\right)}{\mathbb{E}\left[\exp\left(-\frac{1}{\xi} [g(X) + \lambda \cdot f(X)]\right)\right]}.$$

Substituting back into equation (15) gives equation (7).

We can connect these results to our earlier analysis of dual Problem 3.6 by defining an alternative expectation $\widehat{\mathbb{E}}$ using a relative density:

$$\frac{\exp\left[-\frac{1}{\xi}g(X)\right]}{\mathbb{E}\exp\left[-\frac{1}{\xi}g(X)\right]}$$

Then write the objective as

$$\widehat{\mathbb{K}}(\xi; g) \doteq \sup_{\lambda} -\xi \log \widehat{\mathbb{E}} \exp[-\lambda \cdot f(X)] - \xi \log \mathbb{E} \exp\left[-\frac{1}{\xi}g(X)\right].$$

Since the last term does not depend on λ , we may appeal to Theorem 3.8 for the existence of a solution where Restriction 3.7 is imposed under the change of measure.²⁵

A.5 When will the relative entropy constraint bind?

We first give a high-level sufficient condition under which the relative entropy constraint in problem 3.9 binds. Write

$$\mathbb{K}(g; \xi) = \max_{\lambda} -\xi \log \mathbb{E} \left[\exp\left(-\frac{1}{\xi}g(X) + \lambda \cdot f(X)\right) \right] - \xi \kappa.$$

Let $\lambda(g; \xi)$ denote the maximizer in the definition of $\mathbb{K}(g, \xi)$, and define

$$M_1(g; \xi) = \frac{\exp\left[-\frac{1}{\xi}g(X)\right]}{\mathbb{E}\left(\exp\left[-\frac{1}{\xi}g(X)\right]\right)}$$

²⁵For computational purposes, there may be no reason to use the change of measure.

$$M_2(g; \xi) = \frac{\exp \left[-\frac{1}{\xi}g(X) + \lambda(\xi)f(X) \right]}{\mathbb{E} \left(\exp \left[-\frac{1}{\xi}g(X) + \lambda(\xi)f(X) \right] \right)}$$

Restriction A.1.

$$\lim_{\xi \downarrow 0} \mathbb{E} [M_1(g; \xi)g(X)] - \mathbb{E} [M_2(g; \xi)g(X)] > 0$$

Proposition A.2. *Under restriction A.1,*

$$\lim_{\xi \downarrow 0} \frac{\partial}{\partial \xi} \mathbb{K}(g; \xi) = \infty$$

and therefore the relative entropy constraint in problem 3.9 binds for any value of $\kappa > \bar{\kappa}$.

Proof. An application of the Envelope Theorem gives that

$$\begin{aligned} \frac{\partial}{\partial \xi} \mathbb{K}(g; \xi) &= -\log \mathbb{E} \left(\exp \left[-\frac{1}{\xi}g(X) + \lambda(g; \xi) \cdot f(X) \right] \right) - \frac{1}{\xi} \mathbb{E} [M_2(g; \xi)g(X)] - \kappa \\ &= \frac{1}{\xi} \mathbb{H}(g; \xi) - \kappa \end{aligned}$$

where

$$\mathbb{H}(g; \xi) = -\xi \log \mathbb{E} \left(\exp \left[-\frac{1}{\xi}g(X) + \lambda(g; \xi)f(X) \right] \right) - \mathbb{E} [M_2(g; \xi)g(X)].$$

Applying L'Hôpital's rule, we see that

$$\lim_{\xi \downarrow 0} \mathbb{H}(g; \xi) = \lim_{\xi \downarrow 0} \mathbb{E} [M_1(g; \xi)g(X)] - \mathbb{E} [M_2(g; \xi)g(X)] > 0.$$

The result follows. □

Restriction A.1 is difficult to verify in practice. To make things more concrete, we give two somewhat general examples under which the relative entropy constraint will bind.

Example A.3 establishes that the relative entropy constraint will bind in problem 3.9 whenever the target random variable $g(X)$ has a lower bound \underline{g} with arbitrarily small probability near that bound.

Example A.3. *For simplicity, omit the moment condition $\mathbb{E}[Mf(X)] = 0$. Suppose that*

$$(i) \text{ ess inf}[g(X)] = \underline{g} > -\infty,$$

$$(ii) \lim_{\epsilon \rightarrow 0} \mathbf{P} \{g(X) \leq \underline{g} + \epsilon\} = 0,$$

Then for any $\kappa > 0$, the relative entropy constraint in Problem 3.9 will bind.

Example A.3 rules out indicator functions for the choice of g . Bounding such functions may be of interest if the econometrician wishes to consider bounds on distorted probabilities. We consider a version that allows for these in example A.4

Example A.4. *We consider a scalar moment condition with a support condition and consider bounds on indicator functions of the moment function. Suppose*

$$(i) f(X) \text{ is a scalar random variable;}$$

$$(ii) \text{ ess sup}(f(X)) = \mathbf{u} < \infty,$$

$$(iii) \lim_{\epsilon \rightarrow 0} \mathbf{P} \{f(X) \geq \mathbf{u} - \epsilon\} = 0.$$

$$(iv) g(X) = \mathbf{1}_{\{f(X) \geq -r\}} \text{ for } r > 0;$$

Then for any $\kappa > 0$, the relative entropy constraint in Problem 3.9 will bind.

The statement that the relative entropy constraint binds for any $\kappa > 0$ in examples A.3 and A.4 follows immediately from lemmas A.5 and A.6 respectively. These two examples suggest that the relative entropy constraint will bind in many cases of interest even for arbitrarily large choices of κ .

A.6 Auxiliary results

Lemma A.5. *Let $\underline{g} = \text{ess inf } g(X)$ and assume that*

$$\lim_{\epsilon \rightarrow 0} \mathbf{P} \{g(X) \leq \underline{g} + \epsilon\} = 0.$$

Then for any $\kappa > 0$, there exists a constant $\zeta > \underline{g}$ such that for any belief distortion M satisfying $M \geq 0$, $\mathbb{E}[M] = 1$, and $\mathbb{E}[Mg(X)] \leq \zeta$, we must have that $\mathbb{E}[M \log M] > \kappa$.

Proof:

Write

$$h(\epsilon) = \mathbf{P} \{g(X) \leq \underline{g} + \epsilon\}$$

and observe that $h(\epsilon) > 0$ and $h(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. Define an event $A(\epsilon)$ by

$$A(\epsilon) = \{g(X) \leq \underline{g} + \epsilon\}$$

Now, let $\zeta = \underline{g} + \frac{\epsilon}{2}$. Then for any M satisfying the constraints, we have that

$$\begin{aligned} \underline{g} + \frac{\epsilon}{2} &\geq \mathbb{E}[Mg(X)] \\ &= \mathbb{E}[Mg(X)\mathbf{1}_{A(\epsilon)}] + \mathbb{E}[Mg(X)\mathbf{1}_{A(\epsilon)^c}] \\ &\geq \underline{g} \mathbb{E}[M\mathbf{1}_{A(\epsilon)}] + (\underline{g} + \epsilon)\mathbb{E}[M\mathbf{1}_{A(\epsilon)^c}] \\ &\geq \underline{g} + \epsilon \mathbb{E}[M\mathbf{1}_{A(\epsilon)^c}] \\ &= \underline{g} + \epsilon(1 - Q(\epsilon; M)) \end{aligned}$$

where $Q(\epsilon; M) = \mathbb{E}[M\mathbf{1}_{A(\epsilon)}]$. Rearranging, we obtain the bound

$$\frac{1}{2} \geq 1 - Q(\epsilon)$$

which simplifies to

$$Q(\epsilon) \geq \frac{1}{2}.$$

It follows that

$$\mathbb{E}[M|A(\epsilon)] = \frac{\mathbb{E}[M\mathbf{1}_{A(\epsilon)}]}{\mathbb{E}[\mathbf{1}_{A(\epsilon)}]} = \frac{Q(\epsilon)}{h(\epsilon)} \geq \frac{1}{2h(\epsilon)}.$$

Additionally, since $M \geq 0$ we have the trivial inequality

$$\mathbb{E}[M|A(\epsilon)^c] \geq 0.$$

Now, let $\mathcal{F}(\epsilon)$ denote the σ -algebra generated by the event $A(\epsilon)$. Applying Jensen's inequality conditional on $\mathcal{F}(\epsilon)$ to the relative entropy, we obtain

$$\begin{aligned} \mathbb{E}[M \log M] &\geq \mathbb{E}[\mathbb{E}[M|\mathcal{F}(\epsilon)] \log (\mathbb{E}[M|\mathcal{F}(\epsilon)])] \\ &= h(\epsilon) \frac{Q(\epsilon)}{h(\epsilon)} \log \left[\frac{Q(\epsilon)}{h(\epsilon)} \right] + [1 - h(\epsilon)] \left(-\frac{1}{e} \right) \end{aligned}$$

$$\geq \frac{1}{2} \log \left[\frac{1}{2h(\epsilon)} \right] - \frac{1}{e}$$

where the second term comes from the fact that the function $\phi(m) = m \log m$ is bounded from below by $-e^{-1}$. Choosing ϵ sufficiently small so that the lower bound exceeds κ gives the desired result. \square

Lemma A.6. *Let $f(X)$ be a scalar random variable. Assume that $M \geq 0$, $\mathbb{E}[M] = 1$, $\mathbb{E}[Mf(X)] = 0$ and that $\mathbb{P}\{f(X) \leq u\} = 1$. Then for any $r > 0$*

$$\mathbb{E}[M\mathbf{1}\{f(X) \leq -r\}] \leq \frac{u}{u+r}$$

Proof.

$$\begin{aligned} 0 &= \mathbb{E}[Mf(X)] \\ &= \mathbb{E}[Mf(X)\mathbf{1}\{f(X) \leq -r\}] + \mathbb{E}[Mf(X)\mathbf{1}\{f(X) > -r\}] \\ &\leq -r\mathbb{E}[M\mathbf{1}\{f(X) \leq -r\}] + u\mathbb{E}[M\mathbf{1}\{f(X) > -r\}] \\ &\leq -(u+r)\mathbb{E}[M\mathbf{1}\{f(X) \leq -r+u\}]. \end{aligned}$$

Rearranging gives the desired result. \square

Note that this upper bound is sharp so long as X has strictly positive density near \bar{x} and $-r$. It can be approximated by letting M approach a two-point distribution with a point mass at \bar{x} with probability $\pi = \frac{\bar{x}}{\bar{x}+r}$ and a point mass at $-r$ with probability $1 - \pi = \frac{r}{\bar{x}+r}$.

Lemma A.7. *Let $u = \text{ess sup } f(X)$ and assume that*

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}(f(X) \geq u - \epsilon) = 0$$

Then for any $\kappa > 0$ and $r > 0$ such that $\mathbb{P}\{f(X) \leq -r\} > 0$, there exists a constant $\delta > 0$ such that for any belief distortion M satisfying $M \geq 0$, $\mathbb{E}[M] = 1$, $\mathbb{E}[Mf(X)] = 0$ and

$$\mathbb{E}[M\mathbf{1}\{f(X) \leq -r\}] \geq \frac{u}{u+r} - \delta,$$

we must have that $\mathbb{E}[M \log M] > \kappa$.

Proof. Write

$$h(\epsilon) = \mathbb{P}(f(X) \geq u - \epsilon)$$

and observe that $h(\epsilon) > 0$ and $h(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$.

Now, take $\epsilon \in (0, \mathbf{u} + \mathbf{r})$ and define the following events

$$\begin{aligned} A &= \{f(X) \leq -\mathbf{r}\} \\ B(\epsilon) &= \{-\mathbf{r} < f(X) < \mathbf{u} - \epsilon\} \\ S(\epsilon) &= \{f(X) \geq \mathbf{u} - \epsilon\}. \end{aligned}$$

Observe that A , $B(\epsilon)$ and $S(\epsilon)$ are mutually exclusive. Using the fact that $\mathbf{1}_{B(\epsilon)} = 1 - \mathbf{1}_A - \mathbf{1}_{S(\epsilon)}$ with probability one, we obtain

$$\begin{aligned} 0 &= \mathbb{E}[Mf(X)] \\ &= \mathbb{E}[Mf(X)\mathbf{1}_A] + \mathbb{E}[Mf(X)\mathbf{1}_{B(\epsilon)}] + \mathbb{E}[Mf(X)\mathbf{1}_{S(\epsilon)}] \\ &\leq -\mathbf{r}\mathbb{E}[M\mathbf{1}_A] + (\mathbf{u} - \epsilon)\mathbb{E}[M\mathbf{1}_{B(\epsilon)}] + \mathbf{u}\mathbb{E}[M\mathbf{1}_{S(\epsilon)}] \\ &= -\mathbf{r}\mathbb{E}[M\mathbf{1}_A] + (\mathbf{u} - \epsilon)\mathbb{E}[M(1 - \mathbf{1}_A - \mathbf{1}_{S(\epsilon)})] + \mathbf{u}\mathbb{E}[M\mathbf{1}_{S(\epsilon)}] \\ &\leq (\mathbf{u} - \epsilon) - (\mathbf{u} + \mathbf{r} - \epsilon)\mathbb{E}[M\mathbf{1}_A] + \epsilon\mathbb{E}[M\mathbf{1}_{S(\epsilon)}]. \end{aligned}$$

Rearranging, we obtain the lower bound

$$\mathbb{E}[M\mathbf{1}_{S(\epsilon)}] \geq \frac{(\mathbf{u} + \mathbf{r} - \epsilon)}{\epsilon} \left(\mathbb{E}[M\mathbf{1}_A] - \frac{\mathbf{u} - \epsilon}{(\mathbf{u} + \mathbf{r} - \epsilon)} \right)$$

Now, for any M such that

$$\mathbb{E}[M\mathbf{1}_A] \geq \frac{\mathbf{u}}{\mathbf{u} + \mathbf{r}} - \frac{\epsilon}{2} \frac{\mathbf{r}}{(\mathbf{u} + \mathbf{r})(\mathbf{u} + \mathbf{r} - \epsilon)}$$

we have that

$$\begin{aligned} \mathbb{E}[M\mathbf{1}_{S(\epsilon)}] &\geq \frac{(\mathbf{u} + \mathbf{r} - \epsilon)}{\epsilon} \left(\frac{\mathbf{u}}{\mathbf{u} + \mathbf{r}} - \frac{\epsilon}{2} \frac{\mathbf{r}}{(\mathbf{u} + \mathbf{r})(\mathbf{u} + \mathbf{r} - \epsilon)} - \frac{\mathbf{u} - \epsilon}{(\mathbf{u} + \mathbf{r} - \epsilon)} \right) \\ &\geq \frac{(\mathbf{u} + \mathbf{r} - \epsilon)}{\epsilon} \left(\frac{\epsilon}{2} \frac{\mathbf{r}}{(\mathbf{u} + \mathbf{r})(\mathbf{u} + \mathbf{r} - \epsilon)} \right) \\ &\geq \frac{1}{2} \frac{\mathbf{r}}{\mathbf{u} + \mathbf{r}} \end{aligned}$$

It follows that

$$\mathbb{E}[M|S(\epsilon)] = \frac{\mathbb{E}[M\mathbf{1}_{S(\epsilon)}]}{\mathbb{E}[\mathbf{1}_{S(\epsilon)}]} \geq \frac{1}{2h(\epsilon)} \frac{\mathbf{r}}{\mathbf{u} + \mathbf{r}}.$$

Now, let $\mathcal{F}(\epsilon)$ denote the σ -algebra generated by the event $S(\epsilon)$. Applying Jensen's inequality conditional on $\mathcal{F}(\epsilon)$ to the function $\phi(m) = m \log m$, we obtain

$$\begin{aligned} \mathbb{E}[M \log M] &\geq \mathbb{E}[\mathbb{E}[M|\mathcal{F}(\epsilon)] \log (\mathbb{E}[M|\mathcal{F}(\epsilon)])] \\ &\geq h(\epsilon) \frac{\mathbb{E}[M \mathbf{1}_{S(\epsilon)}]}{h(\epsilon)} \log \left(\frac{\mathbb{E}[M \mathbf{1}_{S(\epsilon)}]}{h(\epsilon)} \right) + (1 - h(\epsilon)) \left(-\frac{1}{e} \right) \\ &\geq \frac{1}{2} \frac{r}{u+r} \log \left(\frac{1}{2h(\epsilon)} \frac{r}{u+r} \right) - \frac{1}{e}. \end{aligned}$$

Now choosing ϵ sufficiently small so that the lower bound exceeds κ gives the desired result. \square

B Proofs and derivations for section 5

This appendix expands on results presented in section 5.

B.1 Perron-Frobenius problem

For an arbitrary λ_0

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

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{J}_0 and only well-defined up to a positive scale factor. Notice that

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

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

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

satisfies:

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

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 $(\tilde{\epsilon}, \tilde{\epsilon}_0)$ and $(\hat{\epsilon}, \hat{\epsilon}_0)$ where $\tilde{\epsilon}_0/\hat{\epsilon}_0$ is not constant and $\tilde{\epsilon} \geq \hat{\epsilon}$. Construct the corresponding \tilde{N}_1 and \tilde{M}_T and let \tilde{Q} be a measure-preserving probability consistent with \tilde{N}_1 and is stochastically stable. Since,

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

it follows that

$$\mathbb{E} \left[\tilde{N}_1 \begin{pmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_1 \end{pmatrix} \mid \mathfrak{I}_0 \right] = \begin{pmatrix} \hat{\epsilon} \\ \tilde{\epsilon} \end{pmatrix} \begin{pmatrix} \hat{\epsilon}_0 \\ \tilde{\epsilon}_0 \end{pmatrix}$$

Consider two cases.

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

$$\mathbb{E} \left[\tilde{N}_1 \begin{pmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_1 \end{pmatrix} \mid \mathfrak{I}_0 \right] = \frac{\hat{\epsilon}_0}{\tilde{\epsilon}_0}$$

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

$$\mathbb{E} \left[\tilde{M}_T \begin{pmatrix} \hat{\epsilon}_T \\ \tilde{\epsilon}_T \end{pmatrix} \mid \mathfrak{I}_0 \right] = \frac{\hat{\epsilon}_0}{\tilde{\epsilon}_0}$$

This contradicts the presumption that \tilde{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

$$\mathbb{E} \left[\tilde{N}_1 \begin{pmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_1 \end{pmatrix} \mid \mathfrak{I}_0 \right] = \begin{pmatrix} \hat{\epsilon} \\ \tilde{\epsilon} \end{pmatrix} \begin{pmatrix} \hat{\epsilon}_0 \\ \tilde{\epsilon}_0 \end{pmatrix}$$

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

$$\mathbb{E} \left[\widetilde{M}_T \left(\frac{\hat{e}_T}{\tilde{e}_T} \right) \mid \mathfrak{J}_0 \right] \rightarrow 0$$

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

B.2 Problem solution

As in section 3, 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:

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

Lemma B.2. *Let λ_0^* solve the dual problem for the eigenvalue-eigenfunction pair (ϵ^*, e_0^*) . Moreover, 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

$$1 \leq \liminf_{T \rightarrow \infty} (\epsilon^*)^{-T} \mathbb{E} \left(\exp \left[\sum_{t=1}^T -\frac{1}{\xi} g(X_t) + \hat{\lambda}_{t-1} \cdot f(X_t) \right] \left(\frac{e_T^*}{e_0^*} \right) \mid \mathfrak{J}_0 \right)$$

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

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

Since λ_0^* solves the dual problem,

$$\begin{aligned} 1 &\leq (\epsilon^*)^{-1} \mathbb{E} \left(\exp \left[-\frac{1}{\xi} g(X_1) + \hat{\lambda}_0 \cdot f(X_1) \right] \left(\frac{e_1^*}{e_0^*} \right) \mid \mathfrak{I}_0 \right) \\ &= \mathbb{E} \left(N_1^* \exp \left[(\hat{\lambda}_0 - \lambda_0^*) \cdot f(X_1) \right] \mid \mathfrak{I}_0 \right) \end{aligned} \quad (16)$$

Iterating on inequality (16) for T time periods gives

$$1 \leq \mathbb{E} \left(N_T^* \exp \left[\sum_{t=1}^T (\hat{\lambda}_{t-1} - \lambda_{t-1}^*) \cdot f(X_t) \right] \mid \mathfrak{I}_0 \right).$$

Thus

$$1 \leq \liminf_{T \rightarrow \infty} \mathbb{E} \left(N_T^* \exp \left[\sum_{t=1}^T (\hat{\lambda}_{t-1} - \lambda_{t-1}^*) \cdot f(X_t) \right] \mid \mathfrak{I}_0 \right)$$

with probability one. □

Remark B.3. *Suppose that λ_0^* is essentially unique. Then (16) 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.

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

subject to the constraint:

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

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 B.4. *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 $\mathbb{E}(N_1^* \mid \mathfrak{I}_0) = 1$. Suppose that v_0^* is bounded. Let \hat{N}_1 be some other change of probability measure that induces stochastic stability. Then*

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

Proof. From the dual problem:

$$\mu^* \leq \mathbb{E} \left(\widehat{N}_1 \left[g(X_1) + \xi \log \widehat{N}_1 + \lambda_0^* \cdot f(X_1) + v_1^* + \nu_0^* \right] \mid \mathfrak{I}_0 \right) - v_0^* - \nu_0^* \quad (17)$$

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

$$\mu^* \leq \mathbb{E} \left(\widehat{N}_1 \left[g(X_1) + \xi \log \widehat{N}_1 + v_1^* \right] \mid \mathfrak{I}_0 \right) - v_0^*. \quad (18)$$

Let

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

Iterating on relation (18) gives:

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

Dividing by T and taking limits implies that

$$\begin{aligned} \mu^* &\leq \lim_{T \rightarrow \infty} \mathbb{E} \left(\widehat{M}_T \frac{1}{T} \left[\sum_{t=1}^T g(X_t) + \xi \log \widehat{N}_t \right] \mid \mathfrak{I}_0 \right) + \lim_{T \rightarrow \infty} \frac{1}{T} \left[\mathbb{E} \left(\widehat{M}_T v_T^* \mid \mathfrak{I}_0 \right) - v_0^* \right] \\ &= \int \mathbb{E} \left(\widehat{N}_1 \left[g(X_1) + \xi \log \widehat{N}_1 \right] \mid \mathfrak{I}_0 \right) d\widehat{Q}_0 \end{aligned}$$

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

Remark B.5. 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 B.6. Suppose that equation (17) is satisfied with strictly positive probability. Then we may write

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

where the random variable $b_0 \geq 0$ is strictly positive with positive probability. Provided that

b_0 remains strictly positive with positive probability under the $\hat{\cdot}$ probability measure,

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

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.

B.3 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 $\mathbb{E}[f(X_t) \mid Z_{t-1}] = 0$, 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

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

We could form a test of this restriction by checking if

$$\frac{1}{T} \sum_{t=1}^T \lambda(Z_{t-1}) \cdot f(X_t) \leq -c < 0. \tag{19}$$

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 (19) has positive probability under \mathbf{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 $-c$. Large deviation theory informs us about the limit

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

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 and Varadhan. 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 \mathbf{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 (12). 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

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

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

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

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

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

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.

C Bounding risk premia

To compute the lower and upper bounds on risk premium that we report in section 6, we extend our previous approach as follows.

- set $g(x) = R^w - \zeta R^f$ where ζ is a “multiplier” that we will search over;

- for alternative ζ , deduce $N_1^*(\zeta)$ and $Q_0^*(\zeta)$ as described in the paper;
- compute:

$$\log \int \mathbb{E} [N_1^*(\zeta) R^w \mid \mathcal{I}_0] dQ_0^*(\zeta) - \log \int \mathbb{E} [N_1^*(\zeta) R^f \mid \mathcal{I}_0] dQ_0^*(\zeta)$$

and minimize with respect to ζ ;

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

Two observations:

- i) the objective is not globally convex;
- ii) a similar approach may be used to deduce upper and lower bounds on other functions of moments such as volatility.