PNAS
www.pnas.org

Supplementary Information for

Robust Identification of Investor Beliefs

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

Corresponding author: Lars Peter Hansen.
E-mail: lhansen@uchicago.edu

This PDF file includes:

Tables S1 to S2 (not allowed for Brief Reports)
SI References

www.pnas.org/cgi/doi/10.1073/pnas.2019910117

1. Summary

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

A. Convexity of the divergence. Write:

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

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

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

where

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

Since ϕ is convex,

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

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

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

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

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

B. Perron-Frobenius problem. For an arbitrary λ_0

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

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

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

is positive and has conditional expectation equal to one.

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

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

satisfies:

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

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

Lemma 1.1. *When there are multiple positive eigenvalue solutions for a given λ_0, at most one of them induces a probability measure that is stochastically stable.*

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

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

it follows that

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

Consider two cases.

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

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

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

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

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

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

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

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

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

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

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

Recall the functional equation for the dual problem:

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

Lemma 1.2. *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

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

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

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

Since λ_0^* solves the dual problem,

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

Iterating on inequality Eq. (1) for T time periods gives

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

Thus

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

with probability one. □

Remark 1.3. *Suppose that λ_0^* is essentially unique. Then* Eq. (1) *is satisfied with a strict inequality holding with positive probability. Under geometric ergodicity of process induced by the * probability, the right-hand-side converges to limit that is strictly greater than one.*

Next consider the primal problem.

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

subject to the constraint:

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

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

Lemma 1.4. *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 $\mathbb{E}\left(N_1^* \mid \mathfrak{I}_0\right) = 1$. Suppose that v_0^* is bounded. Let $\hat{N}_1$ be some other change of probability measure that induces stochastic stability. Then*

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

Proof. From the dual problem:

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

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

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

Let

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

Iterating on relation Eq. (3) gives:

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

Dividing by T and taking limits implies that

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

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

Remark 1.5. *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 1.6. *Suppose that equation* Eq. (2) *is satisfied with strictly positive probability. Then we may write*

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

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

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

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

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

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

We could form a test of this restriction by checking if

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

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

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

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

telling us how quickly these probabilities decay to zero.

The initial version of this is the type of large deviation approximation applied to empirical distributions is due to (1) for iid sequences. It has been extended to Markov processes as discussed by (2) and (3). Under some additional regularity conditions, remarkably, the decay can be made to be remarkably close to the minimum relative entropy over the set of possible probability measures relative to P used for representing the evolution of the (X_t, Z_t). In terms of this literature, relative entropy serves as what is called a "rate function".

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

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

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

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

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

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

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

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

- construct $g(x) = g_1(x) - \zeta g_2(x)$ where ζ is a "multiplier" that we will search over;
- for alternative ζ, deduce $N_1^*(\zeta)$ and $\mathsf{Q}_0^*(\zeta)$ as described in the paper;
- compute:

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

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

Observe that the objective is not globally convex.

F. Additional empirical computations. We compare results from relative entropy and conditional quadratic in the table that follows.

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

Table S1. Expected log market return bounds computed with relative entropy and quadratic specifications of conditional divergence. The relative entropy at the minimizing solution for the quadratic divergence is .0319. This is in comparison to .0284 when we use the relative entropy divergence.

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

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

Table S2. Proportional risk compensations computed as $\log \mathbb{E}R - \log \mathbb{E}R^f$ scaled to an annualized percent. See French's data library (link) for definitions of 'small minus big' and 'high minus low'.

References

1. IN Sanov, On the Probability of Large Deviations of Random Magnitudes. *Matiematicheskii Sbornik* **42**, 11–44 (1957).
2. P Dupuis, R Ellis, *A Weak Convergence Approach to the Theory of Large Deviations.* (John Wiley and Sons, Inc., New York), (1997).
3. SR Varadhan, Special Invited Paper: Large Deviations. *The Annals Probab.* **36**, 397–419 (2008).