



## Four types of ignorance: Discussion <sup>☆</sup>

Harald Uhlig <sup>a,b,c,\*</sup>,<sup>1</sup>

<sup>a</sup> Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA

<sup>b</sup> NBER, USA

<sup>c</sup> CEPR, UK



### ARTICLE INFO

#### Article history:

Received 12 December 2014

Accepted 12 December 2014

Available online 19 December 2014

#### Keywords:

Risk

Robustness

### ABSTRACT

This is a comment on Hansen and Sargent (2015). Using a generalized one-period setup, I provide some intuition and perspective.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Let me try to give my own “birds-eye” perspective of this beautiful paper. The authors consider a version of a problem posed by Friedman (1953). A policy maker seeks to minimize a discounted sum of  $E[X_{t+1}^2]$ , where

$$X_{t+1} = \kappa X_t + \beta U_t + \alpha W_{t+1} \quad (1)$$

and where, as benchmark,  $\kappa, \beta, \alpha$  are known parameters, where  $U_t$  is chosen by the policy maker, whereas  $W_{t+1}$  is noise: as a benchmark, it is normally distributed with mean zero and of a given variance. The authors consider four cases of ignorance or distrust of the original model (1). The first follows Friedman (1953), in that the policy maker does not know  $\beta$ , but has a Bayesian prior over  $\beta$ . The other three are examining the problem of a robust decision maker, who is concerned that the outcome  $X_{t+1}$  is particularly bad, due to

2. ... a distribution for  $\beta$  nearby the prior.
3. ... a distribution for  $W_{t+1}$  nearby the benchmark normal.
4. ... a distribution for  $W_{t+1}$  of particular form and nearby the benchmark normal.

After solving the Friedman case, the authors proceed to solve the next three, new cases. They show that concerns about nearby distributions for  $\beta$  give rise to even more caution than the Friedman case, compared to knowing  $\beta$  for sure. Here, the authors examine “multiplier preferences” and “constraint” preferences to constrain how far the decision maker will go in contemplating alternative distributions. The authors provide a lucid and useful discussion into a numerical example illustrating this case. The authors show that concerns about nearby distributions for  $W_{t+1}$  does not give rise to extra

<sup>☆</sup> This research has been supported by the NSF Grant SES-1227280.

\* Correspondence address: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA.

E-mail address: [huhlig@uchicago.edu](mailto:huhlig@uchicago.edu)

<sup>1</sup> I have an ongoing consulting relationship with a Federal Reserve Bank, the Bundesbank and the ECB.

caution, and leaves the decision for  $U_t$  unaltered, compared to the case when  $\beta$  is known. However, specific misspecifications as in the last case reintroduce caution.

This is a beautiful and insightful paper indeed. One can perhaps view it as an introduction to what is possible with the tools provided in the book treatment by Hansen–Sargent (2008), wetting the readers appetite. One can also view it as a paper about the scope and the need for caution for policy makers, when the “true” model is not known: a situation certainly reasonable and appealing to many policy makers out there.

Generally, I found the calculations and discussions about the second case of ignorance (concerns about misspecifying the distribution for  $\beta$ ) the most interesting: I can well imagine that many readers will focus on that case for reading and further research. By contrast, the last case of ignorance is the hardest to digest, and it may well be that future readers will tend to skip that part. That, obviously, would be a shame. The paper is not exactly for the faint of heart, and there occasionally is quite a bit of algebra and notation to digest, before one gets to see the key insights. The key insights are quite a reward for this effort, though, so this is a price well worth paying. In my discussion, I shall try to abbreviate the path to these insights a bit: precision gets lost and the insights get a bit distorted as a result. But every good game needs a cheat sheet, so here it is. Or perhaps, this perspective is even more daunting than the original? In that case, hopefully it is a useful complement.

## 2. A generalized perspective

Let me strip the problem down a bit. There is a quadratic objective function  $h(\cdot)$ . There are some univariate and independent normal probability distributions  $F$  and  $G$ . There are some other univariate probability distributions  $\tilde{F}$  and  $\tilde{G}$  under considerations for the policy maker. There is a notion of distance  $d(\cdot, \cdot)$  between distributions. Here, Hansen and Sargent (2015) use relative entropy (and let us not dwell much here on what one needs to do to make that a proper “distance” and whether and how their results might change as a result).

With these pieces on the table, the policy maker seeks to solve

$$V(x) = \max_u \min_{\tilde{F}, \tilde{G}} \{E[h(y)] + e^{-\delta} V(y)\}$$

$$d(F, \tilde{F}) \leq \eta, d(G, \tilde{G}) \leq \gamma$$

$$y = x + \beta u + \alpha w, \beta \sim \tilde{F}, w \sim \tilde{G}$$

Friedman considered the case  $\gamma = \eta = 0$ . There are two places to introduce concerns for robustness:

1. “Robust-A” sets  $\gamma = 0$ , focussing on robustness concerns and alternative distributions  $\tilde{F}$  regarding the feedback coefficient  $\beta$ .
2. “Robust-B” sets  $\eta = 0$ , focussing on robustness concerns and alternative distributions  $\tilde{G}$  regarding the disturbance term  $w$ .

The use of relative entropy for calculating “distance” has a number of advantages. There is a natural interpretation as a likelihood. One obtains analytical and closed-form results. There is some really beautiful algebra, resulting from it all. But there is also a price to be paid. The algebra can get quite dense. The plots require numerics, anyhow. And one may wonder which results depend specifically on using relative entropy and which results obtain more generally, when thinking about “nearby” probability distributions.

So let me strip down the problem some more, by eliminating the dynamics. So, now the policy maker solves

$$V(x) = \max_u \min_{\tilde{F}, \tilde{G}} \{E[h(y)]\}$$

$$d(F, \tilde{F}) \leq \eta, d(G, \tilde{G}) \leq \gamma$$

$$y = x + \beta u + \alpha w, \beta \sim \tilde{F}, w \sim \tilde{G}$$

It then is quite natural to write down the Lagrangian

$$V(x) = \max_u \min_{\tilde{F}, \tilde{G}} \{E[h(y)] + \theta d(F, \tilde{F}) + \lambda d(G, \tilde{G})\}$$

$$y = x + \beta u + \alpha w, \beta \sim \tilde{F}, w \sim \tilde{G}$$

From this, one can understand the two ways in which Hansen and Sargent approach solving for the robust solutions:

1. For “constraint preferences” or the operator  $\mathbf{C}^2$ : fix  $\eta, \gamma$ . Then,  $\theta, \lambda$  depend on  $(x, u)$ .
2. For “multiplier preferences” or the operator  $\mathbf{T}^2$ : fix  $\theta, \lambda$ . Then,  $\eta, \gamma$  depend on  $(x, u)$ .

For their last case, they also consider what I shall call “Robust-C”, where  $\lambda = \lambda(x, u)$  for some specific function of  $x, u$ .

Let me analyze this a bit further. Let me first substitute out  $y$  and state the Lagrangian problem as

$$\max_u \min_{\tilde{F}, \tilde{G}} E_{\tilde{F}, \tilde{G}}[h(x + \beta u + \alpha w)] + \theta d(F, \tilde{F}) - \lambda d(G, \tilde{G}) \quad (2)$$

where  $\beta \sim \tilde{F}$ ,  $w \sim \tilde{G}$ . Given  $x$ , suppose, the solution is  $\bar{u}, \bar{F}, \bar{G}$ . What does one typically do, when faced with a (differentiable) maximization problem? Yes, let us calculate first-order conditions.

Doing so for probability distributions may feel a bit strange, but it is not that hard. To warm up, consider the case where the  $F$ 's are probability distributions over finitely many given points  $\beta_1, \dots, \beta_n$ . Then, some  $F$  can be represented as a vector in  $\mathbb{R}^n$  with the property that each vector entry is between 0 and 1 and that the vector entries sum to unity. Let us consider a function  $f(F)$  from such probability vectors to the real line. Taking the derivative now just means taking the derivative of a function defined<sup>2</sup> on  $\mathbb{R}^n$ . The derivative, evaluated at some  $\bar{F}$ , is a gradient or “row” vector  $v = f'(\bar{F})$  in  $\mathbb{R}^n$ . One can think of that as a linear mapping, telling us, what it will do to some other vector  $z \in \mathbb{R}^n$ , per calculating the inner product  $v(z) = v \cdot z$ . So, for some other distribution  $\hat{F}$  and with  $z = \hat{F} - \bar{F}$ , this is the term that shows up in the usual first-order linear approximation

$$f(\hat{F}) - f(\bar{F}) \approx v \cdot (\hat{F} - \bar{F})$$

This is basic calculus, of course. And it is also basic calculus, that this approximation is exact, if  $f(\cdot)$  happens to be a linear or affine-linear function. And finally, the first-order condition of setting that derivative to zero is the same<sup>3</sup> as stating that the inner product  $v \cdot (\hat{F} - \bar{F})$  is zero, for all probability vectors  $\hat{F}$ .

For more general probability distributions, these derivatives turn into Fréchet derivatives: once again, they are linear mappings. The first-order conditions now say, that for all  $\hat{F}, \hat{G}$

$$\frac{\partial}{\partial u}: 0 = E_{\bar{F}, \bar{G}}[\beta h'(x + \beta u + \alpha w)] \quad (3)$$

$$\frac{\partial}{\partial \tilde{F}}: 0 = E_{\hat{F} - \bar{F}, \bar{G}}[h(x + \beta u + \alpha w)] + \theta d_2(F, \bar{F})(\hat{F} - \bar{F}) \quad (4)$$

$$\frac{\partial}{\partial \tilde{G}}: 0 = E_{\bar{F}, \hat{G} - \bar{G}}[h(x + \beta u + \alpha w)] + \gamma d_2(G, \bar{G})(\hat{G} - \bar{G}) \quad (5)$$

where I have exploited the linearity of the expectation operator, as a function of the probability distribution and where, say,  $E_{\hat{F} - \bar{F}, \bar{G}}[h(y)]$  stands for the difference between the expectation of  $h(y)$  once calculated using the distribution  $\hat{F}$  and once calculated using the distribution  $\bar{F}$ .

### 3. Robust-A: intuition

One can use this generalized perspective to gain some intuition for the “Robust-A” case and the results in Figs. 1 and 4 in Hansen and Sargent. Start with the constraint preferences case. A given amount of distance in the distribution for  $\beta$  is now permitted. Graphically, this is presented in Fig. 1 as some variation for the outcome  $x + \beta u$ , indicated by the dotted vertical lines. Drawn this way, it is hard to distinguish between Bayesian uncertainty regarding  $\beta$  or robustness concerns regarding the distribution for  $\beta$ : perhaps then it is not surprising that both deliver rather similar decision rules indeed, as Fig. 4 in Hansen and Sargent shows. Examine now the resulting variation in the objective function  $h(x + \beta u + w)$ . It is larger for larger  $x$ : this is obvious for  $u = 0$ , and it still holds generally, if the decision rule  $u(x)$  does not overcompensate for  $x$ , as is plausible. The larger resulting variation in the objective function  $h$  translates into a larger potential loss: this is the “fear” expressed in the robust-control max–min calculus of always having to play against the worst distribution. The larger potential loss means a larger shadow value  $\theta$  on constraining the original variation in  $\beta$  or a larger Lagrange multiplier for that constraint for larger  $x$ . This is the result in Fig. 4 in Hansen–Sargent, and I suspect that it carries over to this more general perspective.

One can now turn this perspective on its head and enforce a multiplier  $\theta$  of the same size or the same utility costs for misspecifications regarding the  $\beta$ -distribution, as is done in the multiplier preference case. In Fig. 2, this is shown as imposing the same range of variation for the objective function  $h$ , which now translates into a much narrower range for  $x + \beta u$ , if  $x$  is larger. Unless the original distribution  $F$  is a point mass, this now translates into a value for  $u$  closer to zero. This is the result in Fig. 1 of Hansen and Sargent for larger  $x$ . It is offset for small values of  $x$  by the optimal linear reaction to  $x$ , when  $\beta$  is known: there, the objective function  $h$  is sufficiently flat, so that the force of narrowing the range for  $x + \beta u$  in Fig. 2 is too small by comparison. Hansen and Sargent in their Fig. 1 show a hump-shaped decision rule for  $u$  as a function of  $x > 0$  (and the same for  $x < 0$ ): approximately linear in  $x$  for small  $x$ , but converging to zero for larger  $x$ . The arguments here suggest that once again this insight may be considerably more general (possibly allowing for multiple humps), but I have not tried to work out the algebra. I suspect that one can make considerable progress here by investigating the implications of the first-order conditions stated above or, potentially, considering second-order considerations.

<sup>2</sup> Strictly speaking, it may only be defined on the simplex of probability vectors.

<sup>3</sup> There is a subtle point here, in that we only insist on the derivative being zero in the “direction” of other probability distributions. Indeed, we might have been “lazy” and define  $f$  only the probability vectors simplex.

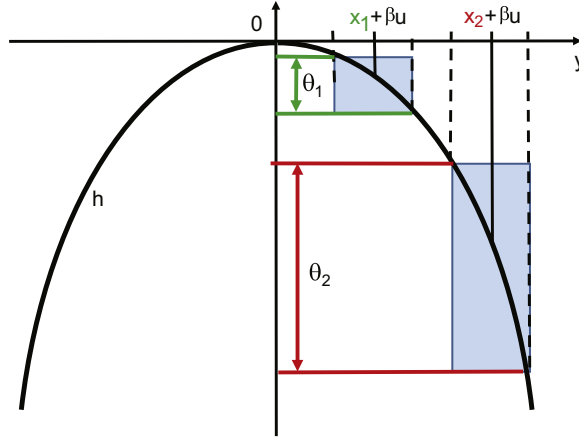


Fig. 1. Intuition for Fig. 4 in Hansen–Sargent and the Robust-A constraint-preferences case.

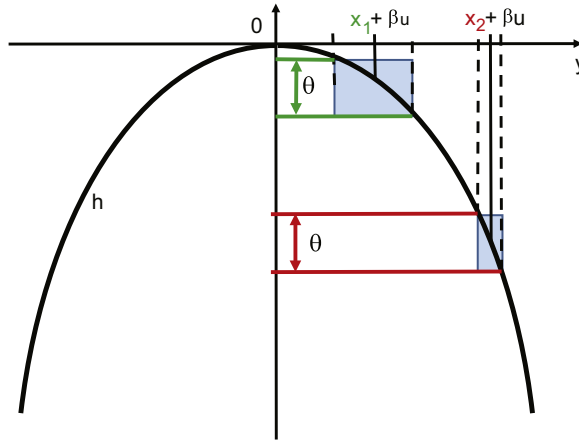


Fig. 2. Intuition for Fig. 1 in Hansen–Sargent and the Robust-A multiplier-preferences case.

**4. Robust-B: the no-increase-in-caution result**

Consider now the “Robust-B” problem of focussing on robustness concerns and alternative distributions  $\tilde{G}$  regarding the disturbance term  $w$ . Hansen and Sargent show the surprising and negative “no-increase-in-caution” result in Section 5, that such robustness concerns do not alter the policy choice  $u$ . Can we get some insight into that result, using the generalized perspective above?

To that end, examine the first-order condition (4) with respect to  $\tilde{G}$ . Suppose that  $\hat{G}$  differs from  $\bar{G}$  by a small shift  $\epsilon$  in  $w$ -space. Indeed, the shift in the mean of the normal distributions considered by Hansen and Sargent is exactly of this kind. Calculate

$$\begin{aligned} & E_{\bar{F}, \hat{G}}[h(x + \beta u + \alpha w)] \\ &= E_{\bar{F}, \bar{G}}[h(x + \beta u + \alpha(w + \epsilon))] \\ &\approx E_{\bar{F}, \bar{G}}[h(x + \beta u + \alpha w)] + \epsilon E_{\bar{F}, \bar{G}}[h'(x + \beta u + \alpha w)] \end{aligned}$$

Recall the first FOC (3). Let us additionally assume that  $\bar{F}$  is a pointmass, i.e., that  $\beta$  is known with certainty, and additionally, that it is different from zero. Then, (3) implies  $E_{\bar{F}, \bar{G}}[h'(x + \beta u + \alpha w)] = 0$ . Together with the calculation above, this implies

$$E_{\bar{F}, \hat{G} - \bar{G}}[h(x + \beta u + \alpha w)] = 0$$

Plug into the third first-order condition (5): the first term drops out and what remains is

$$0 = \gamma d_2(G, \bar{G})(\hat{G} - \bar{G})$$

The only way to satisfy this equation for a proper distance measure is per  $\hat{G} - \bar{G} = 0$ . Put differently, adding some extra shift  $\epsilon$  only “costs distance” and it is better avoided. Allowing that extra flexibility of picking an  $\epsilon$ -shifted  $\hat{G}$  does not yield any

advantages to the “evil probability-distribution-picking agent. And therefore, the original solution  $\bar{u}$  is still the solution in this “Robust-B” problem.

It seems to me that this may be what is truly at heart and behind the “no-increase-in-caution” result in Hansen–Sargent. This somewhat heuristic reasoning shows that one may obtain this result rather quickly from examining the basic first-order conditions, with little additional algebra.

## 5. Caution versus petrification

The algebra for the “Robust-B” case also shows the limits of the “no-increase-in-caution” result. If the  $\hat{G}$ -alternatives were something else than shifts of the entire distribution by some  $\epsilon$ , it does not seem likely that the “no-increase-in-caution” result still obtains. In fact, the “no-increase-in-caution” result might be something like a benchmark in the middle. Depending on the type of  $\hat{G}$ -alternatives, one might get an increase in caution: this is, in essence, what Hansen and Sargent strive for in their analysis of the Robust-C case. But one might just as well get the opposite, and a decrease in caution – or better, a greater nervousness by the policy maker to do something in order to avoid disaster. Sometimes, the worst thing is to do nothing at all, and the most cautious course of action is to act decisively. When a truck is coming straight at you on a narrow road, veer to the left or veer to the right. Do not keep heading straight for the truck.

It is not hard to come up with real-life policy scenarios, where there were was a strong desire by policy makers to do something: whether rightfully so would require a full analysis. Examples are the 2008 financial crisis, the 2010 European debt crisis, the Ebola outbreak in 2014 and so forth. One fascinating possibility with the framework provided to us by Hansen and Sargent is to shed light on when caution implies decisive action. It seems to me that this possibility already arises in the simple case discussed here.

## 6. A Bayesian interpretation

Notice that the first-order conditions above are same, no matter whether the problem is stated as  $\max_u \min_{\hat{F}, \hat{G}}$  or as  $\min_{\hat{F}, \hat{G}} \max_u$ . This is familiar from standard calculus. But while the first way of stating the problem is a robust control decision problem, the second way of stating the problem becomes a Bayesian decision problem for  $u$ , given the selected distributions of the “outer”  $\min_{\hat{F}, \hat{G}}$  step. So, the key for re-interpreting the problem in a Bayesian fashion may be whether it is indeed legitimate to characterize the optimum, using first-order conditions only or whether the sequence of optimization matters. Standard calculus reasoning may go a long way to resolve this in any specific formulation.

## 7. Entropy, catastrophes and policy objectives

In my discussion, I have used a general distance measure to evaluate “nearby” distributions. Hansen and Sargent argue that relative entropy is particularly attractive here. The key argument is that policy makers presumably wish to guard against alternatives that are hard to detect econometrically. It is important to recognize, however, that there is no a priori reason for this principle. It may be fine for some applications. It may not be fine for others.

It is plausible that policy makers do not care so much about fine-tuning for some small misspecification of their model at hand, but really care about guarding against catastrophes. Central bankers like to avoid the specter of deflation. Policy makers generally like to avoid the meltdown of economies. I buy into the Hansen–Sargent perspective that policy makers do not care so much about getting it exactly right as they care about not getting it exactly wrong. But “not getting it exactly wrong” may be about some catastrophic scenario, whether easy to detect or not. Moreover, it is fine to ascribe some particular objective function to policy makers for the purpose of analysis: in practice, they might be hard-pressed to be specific enough about it, and that may matter too. In any case, the objective of the policy maker ought to matter for calculating “distance”. Let me illustrate these points with two abstract examples, in the language of the Hansen–Sargent formulation.

For the first example, suppose that  $F$  and the alternative  $\tilde{F}$  both are distributions on  $\beta \in \{1, 1000\}$ , assigning probability weights  $\{1 - \epsilon, \epsilon\}$  and  $\{1 - \tilde{\epsilon}, \tilde{\epsilon}\}$  respectively, where  $0 < \epsilon < \tilde{\epsilon} < 1$ . Indeed, suppose that both  $\epsilon$  are small, but that  $\tilde{\epsilon}$  is vastly smaller than  $\epsilon$ :  $\epsilon = e^{-10,000}$  and  $\tilde{\epsilon} = e^{-6} \approx 1/400$  will do. The likelihood ratio  $m(\beta)$  is given by the two numbers  $\{(1 - \tilde{\epsilon})/(1 - \epsilon), \tilde{\epsilon}/\epsilon\}$  or approximately  $\{1, e^{9994}\}$  for the specific numbers given. The relative entropy  $\text{ent}(\tilde{F}, F)$  is

$$\text{ent}(\tilde{F}, F) = (1 - \tilde{\epsilon}) \log \left( \frac{1 - \tilde{\epsilon}}{1 - \epsilon} \right) + \tilde{\epsilon} \log \left( \frac{\tilde{\epsilon}}{\epsilon} \right) \approx 25$$

for the specific example. For relative entropies, that is pretty far. For comparison: for normal densities, the mean would need to be shifted by about seven standard deviations to create that kind of an entropy-distance. Pretty far indeed.

So, a Hansen–Sargent policy maker should not be concerned about the alternative  $\tilde{F}$ . Essentially, if  $\beta = 1000$  gets realized in a short sample, the policy maker can be quite sure that it is really the alternative distribution describing the uncertainty regarding  $\beta$ , it is an easily distinguishable probability distribution.

But consider such a policy maker. Perhaps, the policy maker typically takes some action  $U_t$  rather different from zero, which then is justified by a “normal times” draw of  $\beta = 0$ : perhaps this has happened hundred times in a row and it should happen many more times under  $F$ , and there is no good reason to worry about  $\beta = 1000$  under  $F$ . But then suddenly

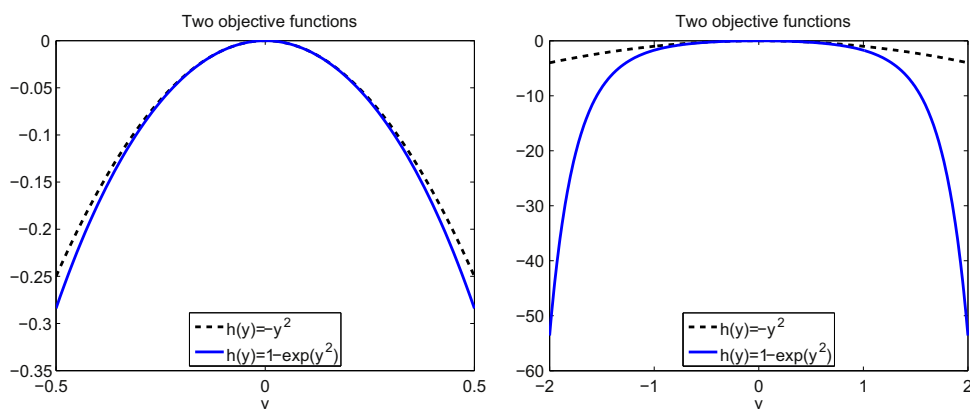


Fig. 3. Comparing two objective functions.

$\beta = 1000$  gets drawn. The policy maker now knows that, very likely,  $\tilde{F}$  and not  $F$  is at work. But by now, it is too late: if some action  $U_t$  reasonably far from zero has already been taken, this creates a huge loss per  $X_{t+1} = X_t + 1000 * U_t + \alpha W_{t+1}$ . It seems to me that the policy maker would have been well advised to be cautious against the possibility  $\tilde{F}$ , even though it is far away in an entropy sense.

It may well be that this is a small sample problem. If the policy maker had a reasonably long sample of  $\beta$ -draws before embarking on policy choices  $U_t$ , there would have been no problem in distinguishing  $F$  from  $\tilde{F}$ . But is not that often precisely the problem policy makers face? Their samples are small, and there is a rich amount of data to consider. Tail probabilities for rare events and catastrophes matter, and policy makers do not have the luxury to wait for a long time to decide whether these tail probabilities are small or truly tiny: “small” may be too large.

For the second example and regarding the point that the objective function matters for calculating distance, suppose the objective function is  $E[1 - e^{y^2}]$  rather than  $E[-y^2]$ . Fig. 3 compares these two objective functions: they nearly coincide for moderate values of  $y$ , but with the new objective function, large values for  $|y|$  are far worse than for the quadratic objective. Now, if  $W_{t+1}$  and thus  $X_{t+1}$  are normally distributed, one may go ahead and calculate the optimal choice for  $U_t$ , just as before. But suppose that  $W_{t+1}$  and thus  $X_{t+1}$  are  $t_{1,000,000}$ -distributed, as the alternative density. As a result,  $V = -\infty$ , no matter which  $U_t$  is chosen: it is pointless now for the policy maker to put much effort into picking a best  $U_t$ . More generally, it would not be hard to modify this example a bit and construct “intermediate” cases, where the prediction regarding the optimal  $U_t$  under the alternative become pretty wild. So, should not robustness in light of the objective be on the table? What, if the policy maker actually is unsure about the objective function? Should perhaps decision rules be sought that are robust to misspecifications of the objective function?

## 8. Conclusion

Hansen and Sargent keep paving the way for thinking about the very practical issue of model misspecification and the search for robust decision rules, see also Uhlig (2012). This one here is a nifty, little paper, exposing key elements in a small-scale model. There are lots of insights, lots of food for thought, lots of details to chew on.

I do buy into their perspective that policy makers typically seek to find decision rules that do robustly well, even if the underlying model has been somewhat misspecified. It is very attractive and plausible from a practical perspective to adopt the Hansen–Sargent perspective that policy makers do not care so much about getting it exactly right as they care about not getting it too wrong. Hansen and Sargent push us to take this very practical problem, and think about it using the theoretical tools of research economists. They discipline their exercise by considering alternatives that are “nearby” in an econometric sense: essentially, they wish to be robust against alternatives that are hard to distinguish, when using likelihood ratio tests. Hansen and Sargent reach the conclusion for their particular set-up that this translates into reacting even less to circumstances.

I believe, though, that it is important to recognize the limitations of this particular conclusion. It may well be that the cautious thing is to act decisively: indeed this may already arise in their framework, with small modifications. And it may well be that the misspecification alternatives, which policy makers need to worry about, are very clearly visible and easy to distinguish econometrically, once they manifest themselves. It is not knowing whether and when they occur, which may be the important challenge to policy makers. But all these are matters of debate that merit closer attention in future research.

Overall, this is a powerful perspective and paradigm, and a fascinating research agenda. I wish more researchers would take it up. Hansen and Sargent already have done deep and powerful work exploring big stretches of the coastal areas of this continent. But a large continent it is. A large terra incognita still awaits us. For those who dare venture into this unknown, I recommend taking the Hansen–Sargent charts and this particular paper along the way: it will serve them well.

## References

- Friedman, Milton, 1953. The effects of full employment policy on economic stability: a formal analysis. In: Friedman, Milton (Ed.), *Essays in Positive Economics*, University of Chicago Press, Chicago, IL, pp. 117–132.
- Hansen, Lars Peter, Sargent, Thomas J., 2008. *Robustness*. Princeton University Press, Princeton, NJ.
- Hansen, Lars Peter, Thomas J. Sargent, 2015. Four types of ignorance. *J. Monet. Econ.* 69, 97–113, <http://dx.doi.org/10.1016/j.jmoneco.2014.12.008>.
- Uhlig, Harald, 2012. Agents as Empirical Macroeconomists: Thomas J. Sargent's Contribution to Economics. *Scand. J. of Economics* 114 (4), 1055–1081.