A robust Hansen–Sargent prediction formula

Kenneth Kasa*

Research Department, Federal Reserve Bank of San Francisco, P.O. Box 7702, San Francisco, CA 94120, USA

Abstract

This paper derives a formula for the optimal forecast of a discounted sum of future values of a random variable, where optimal is defined in terms of the minimized $H^\infty$-norm of the forecast error. This problem reflects a preference for robustness in the presence of (unstructured) model uncertainty. The paper shows that revisions of a robust forecast are more sensitive to new information, and discusses the relevance of this result to previous findings of excess sensitivity of consumption and asset prices to new information. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Robust control; Model uncertainty

JEL classification: C44; C63

1. Introduction

The problem of forecasting a discounted sum of the future values of a random variable arises often in dynamic economic models. For example, it is central to Permanent Income models of consumption and to Present Value models of asset pricing. In a classic paper, Hansen and Sargent (1980,1981a) used Wiener-Kolmogorov prediction theory to derive a convenient closed-form expression for the solution of this forecasting problem. The simplicity and generality of their formula has made it a key input in the analysis and econometric evaluation of dynamic rational expectations models. ¹

When deriving their formula Hansen and Sargent assume the decision-maker wants to minimize the mean-squared error of his forecast. This metric is so commonly applied that its use often goes unquestioned. It is important to keep in mind, however, that its rigorous justification rests on fairly restrictive assumptions; namely, linear-quadratic models with complete information about system dynamics.

¹The results in Hansen and Sargent (1980,1981a) apply to the scalar case. In Hansen and Sargent (1981b) the results are extended to the vector case.
This paper retains the linearity assumption while backing away from the assumption of complete information about the model’s dynamics. Instead, I adopt the perspective of the emerging literature on ‘robust control’, which regards a model’s disturbance process as capturing a complex amalgam of potential model misspecifications, e.g., omitted variables, neglected nonlinearities, or misspecified statistical distributions. To ensure robustness against such a broad spectrum of potential misspecifications, the decision-maker resorts to a minmax objective function. With unstructured model uncertainty it makes sense to use a policy that guarantees adequate performance under a broad spectrum of circumstances, even if this means sacrificing some performance if in fact the hypothetical model is correctly specified. The way to guarantee adequate performance is to choose a policy that minimizes the maximum error. Any other objective implicitly commits you to a probability weighting across potential misspecifications, which is counter to the whole spirit of robust control. After all, what if your assessment of these probabilities is itself misspecified?\(^2\)

Applied to the present value forecasting problem considered in this paper, the notion of a robust forecast is based on uncertainty about the stochastic properties of the process you are trying to forecast. For example, Hansen and Sargent (1980,1981a) derive their formula assuming the forecaster knows the Wold representation of the process. In contrast, this paper admits the possibility that this Wold representation is misspecified. A robust present value forecast minimizes the maximum forecast error across all potential misspecifications, where the only restriction placed on the misspecifications is that they be norm bounded. It turns out that robust present value forecasts are more sensitive to new information. At the end of the paper I offer a few conjectures about the relevance of this result to previous findings of ‘excess sensitivity’ of consumption and asset prices to new information.

2. The Hansen–Sargent prediction formula

Many economic models require the solution of the following problem:

\[
y_t = E_t \sum_{j=0}^{\infty} \lambda^j x_{t+j}
\]

where \(y_t\) is a decision variable and \(x_t\) is an exogenous forcing process. For example, in the simplest Permanent Income models of consumption, \(y_t\) would be consumption and \(x_t\) would be income. Alternatively, in asset pricing models \(y_t\) could be the price of a stock or long-term bond, and \(x_t\) would then be dividends or short-term interest rates. The defining characteristic of this class of models is the assumption that the discount factor, \(\lambda\), is constant. This is a highly restrictive theoretical assumption, and given the rejections during the 1980s of models based on Eq. (1), most subsequent analysis has been based on relaxing this assumption in one way or another.

As noted earlier, Hansen and Sargent assume the decision-maker knows the Wold representation of \(x_t\). Given this assumption it makes sense to minimize the mean-squared error of the forecast. Letting

\(^2\)Note that this distrust of prior probabilities is related to the distinction between ‘risk’ and (Knightian) ‘uncertainty’, where uncertainty is defined to be a situation where the Savage axioms (typically the ‘Sure Thing Principle’) do not apply. See Hansen et al. (1999) for further motivation of robust control.
the Wold representation be $x_t = d(L)e_t$, we have then the following frequency domain representation of the forecasting problem:

$$
\min_{f(z) \in H^2} \frac{1}{2\pi i} \oint \frac{d(z)}{1 - \lambda z^{-1}} - f(z) \left\| \frac{dz}{z} \right\|^2
$$

(2)

where $\oint$ denotes contour integration around the unit circle, and $H^2$ denotes the Hardy space of square-integrable analytic functions on the unit disk. The time-domain solution for $y_t$ is then given by $y_t = f(L)\epsilon_t$. Restricting the $z$-transform of $f(L)$ to lie in $H^2$ guarantees the forecast is 'causal', i.e., based on a square-summable linear combination of current and past values of the underlying shocks, $\epsilon_t$.

The solution of the optimization problem in (2) is a standard result in Hilbert space theory (see, e.g., Young (1988, p. 188). It is given by:

$$
f(z) = \left[ \frac{d(z)}{1 - \lambda z^{-1}} \right]_+
$$

(3)

where $[\cdot]_+$ denotes an `annihilation operator', meaning 'ignore negative powers of $z$'. Given (3), Hansen and Sargent (1980,1981a, Appendix A) use results from complex analysis to derive the following convenient expression for (3):

$$
f(z) = \frac{zd(z) - \lambda d(\lambda)}{z - \lambda}
$$

(4)

Note that $f(z) \in H^2$ by construction, since the pole at $z = \lambda$ is cancelled by a zero at $z = \lambda$. Eq. (4) neatly summarizes the cross-equation restrictions that are the hallmark of rational expectations models. Applications of this equation are simply too numerous to mention.

3. A robust Hansen–Sargent prediction formula

The problem addressed in this section is the following: What if the decision-maker does not know the process generating $x_t$, but still has to evaluate the forecasting problem in (1)? In this case, (3) could produce very poor results. For example, suppose that the $z$-transform of the actual process is $d''(z) = d'(z) + \Delta(z)$, where $d''(z)$ is the original benchmark, or nominal model, and $\Delta(z)$ is a (one-sided) perturbation function. Applying (3) in this case yields the following mean-squared forecast error:

$$
\mathcal{L}^a = \mathcal{L}^n + \|\Delta(z)\|^2 + \frac{2}{2\pi i} \oint \left\| \Delta(z) \right\|^2 \left\| \frac{dz}{z} \right\|^2

\frac{d(z)}{1 - \lambda z^{-1}} \frac{dz}{z}
$$

(5)

where $[\cdot]_-$ denotes an annihilation operator that retains only negative powers of $z$, and $\mathcal{L}^a$ and $\mathcal{L}^n$ denote actual and nominal mean-squared forecast errors. Clearly, $\mathcal{L}^a$ could be much greater than $\mathcal{L}^n$, ...
even when $\|\Delta(z)\|_2^2$ is small, depending on how $\Delta(z)$ interacts with $\lambda$ and $d(z)$. Specifically, applying Cauchy’s Residue Theorem to (5) yields:

$$\mathcal{L}^a = \mathcal{L}^\infty + \|\Delta(z)\|_2^2 + 2d(\lambda)[\Delta(\lambda) - \Delta(0)]$$

(6)

Notice that the last term in (6) is scaled-up $d(\lambda)$ which, though bounded, could be quite large. Hence, it is desirable to construct a forecast with greater robustness to model misspecification.

This greater robustness can be achieved simply by switching norms. Rather than evaluate forecast errors in the $H^2$ sum-of-squares norm, we are now going to evaluate them in the $H^\infty$ supremum norm. In the $H^\infty$-norm the optimal forecasting problem becomes:

$$\min_{f(z)\in H^\infty} \max_{|z|=1} \left| \frac{d(z)}{1 - \lambda z^{-1}} - f(z) \right|^2$$

(7)

where $H^\infty$ denotes the Hardy space of essentially bounded analytic functions on the unit disk.

The solution of the optimization problem in (7) is contained in the following proposition:

**Proposition.** The solution of the robust Present Value forecasting problem in Eq. (7) is unique, and is given by:

$$f(z) = \frac{zd(z) - \lambda d(\lambda)}{z - \lambda} + \frac{\lambda^2}{1 - \lambda^2} d(\lambda)$$

(8)

The proof of this proposition is by construction. The key step is to notice that (7) is an example of Nehari’s Approximation Problem, i.e., minimize the $H^\infty$ distance between a two-sided $L^\infty$ function and a one-sided $H^\infty$ function. Nehari provided a general existence and uniqueness theorem for this problem, which relates the solution to the Hankel norm of the $L^\infty$ function, i.e., $d(z)/(1 - \lambda z^{-1})$ in our case (see, e.g., Young (1988, p. 190)).

By following Duren (1970, pp. 136–142), one can actually construct the solution by converting (7) to an equivalent minimum norm interpolation problem. By assumption, $d(z)$ is one-sided, so the target $L^\infty$ function in (7) has a single pole at $z = \lambda$. With this pole, construct the Blaschke factor, $(z - \lambda)/(1 - \lambda z)$. Since Blaschke factors have a modulus of unity on the unit circle, we can re-write (7) as follows:

$$\min_{f(z)\in H^\infty} \|T(z) - B(z)f(z)\|_\infty$$

(9)

where

$$T(z) = B(z) d(z)(1 - \lambda z^{-1})^{-1} = zd(z)(1 - \lambda z)^{-1}$$

(10)

Now set

$$f(z) = (T(z) - \varphi(z))/B(z)$$

(11)

In general, we should write ‘inf’ and ‘sup’ in (7), but in our case it turns out that the extrema are attained.

See Kasa (1999) for an example of this solution method in the context of robust optimal stabilization policy.
Notice that if \( \varphi(z) \) satisfies the interpolation condition, \( \varphi(\lambda) = T(\lambda) \), then \( f(z) \in H^\infty \), since the zero in the denominator at \( z = \lambda \) will be cancelled by a zero in the numerator at \( z = \lambda \). Plugging this candidate \( f(z) \) into (9) gives the new problem:

\[
\min_{\varphi(z) \in H^\infty} \| \varphi(z) \|_\infty \quad \text{s.t.} \quad \varphi(\lambda) = T(\lambda)
\]

(12)

It turns out that with a single pole the solution for \( \varphi(z) \) is just a constant. Thus, to satisfy the interpolation condition in (12) we have:

\[
\varphi(z) = \lambda d(\lambda)(1 - \lambda^2)^{-1}
\]

(13)

Finally, plugging (13) into (11) gives (8).

It is clear from comparing Eqs. (4) and (8) that revisions of a robust Present Value forecast are more sensitive to new information. This is of some interest given past rejections of traditional Present Value models of consumption and asset pricing, which abstract from model uncertainty. For example, Shiller (1981) found that stock returns are far more volatile than revisions in expected future dividends. Similarly, Flavin (1981) showed that consumption is more volatile than expected future income. Subsequent work has shown that a time-varying discount factor or alternative de-trending procedures can partially explain these ‘excess sensitivity’ findings. This paper suggests that model uncertainty and a preference for robustness might also play a role.

4. Conclusion

Linear-quadratic control methods are powerful because they exploit Certainty Equivalence, which allows a separation between prediction and control. The problem solved in this paper involves an unconventional mixing of standard \( H^2 \)-control results with newer results from the literature on \( H^\infty \)-control. Specifically, this paper derives a robust \( H^\infty \) solution to the prediction part of a traditional \( H^2 \)-control policy. The decision-theoretic foundations of this mixture have yet to be investigated. It is of course possible to adopt a robust approach to both the prediction and control parts of the decision rule. For example, Hansen et al. (1999) solve a robust version of a general equilibrium Permanent Income model. They show that model uncertainty can indeed help to explain various consumption and asset pricing puzzles. However, standard dynamic programming approaches to robust control problems do not distinguish clearly between robust control and robust prediction. Hence, the robust prediction problem solved here is of independent interest.

References