European Journal of Operational Research, 197: 203–213 (2009).

Info-Gap Forecasting and the Advantage of Sub-Optimal Models

Yakov Ben-Haim
Yitzhak Moda'i Chair in Technology and Economics
Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000 Israel
Tel: +972-4-829-3262. Fax: +972-4-829-5711
yakov@technion.ac.il

Contents

Abstract

We consider forecasting in systems whose underlying laws are uncertain, while contextual information suggests that future system properties will differ from the past. We consider linear discrete-time systems, and use a non-probabilistic info-gap model to represent uncertainty in the future transition matrix. The forecaster desires the average forecast of a specific state variable to be within a specified interval around the correct value. Traditionally, forecasting uses a model with optimal fidelity to historical data. However, since structural changes are anticipated, this is a poor strategy. Our first theorem asserts the existence, and indicates the construction, of forecasting models with sub-optimal fidelity to historical data which are more robust to model error than the historically optimal model. Our second theorem identifies conditions in which the probability of forecast success increases with increasing robustness to model error. The proposed methodology identifies reliable forecasting models for systems whose trajectories evolve with Knightian uncertainty for structural change over time. We consider various examples, including forecasting European Central Bank interest rates following 9/11.

Keywords. Forecasting, decision support, info-gaps, robustness, model uncertainty.

[0]\papers\forecast\vss09.tex 25.3.2008.

Processing (MDS)

93B35 Sensitivity (robustness), 93B51 Design techniques (robust design), 93E10 Estimation and detection.

1 The Challenge of Forecasting

A popular Danish saying asserts that "prediction is always difficult, especially of the future". Variation is the only constant: "as they step into the same rivers, different and still different waters flow" as Heraclitus is reported to have said. Both the importance and the difficulty of forecasting result from things changing.

Planning and preparedness for an avian flu pandemic has been a major concern recently. "The nub of the problem lies in the inherent variability of the virus, which makes prediction difficult." (Smith, 2006, p.392). In space travel, accurately predicting the genetic damage from cosmic radiation on the long trip to Mars is difficult due to incomplete understanding of mechanisms by which cells might be able to repair accumulated damage (Parker, 2006).

In econometrics, "the [data] generating process is unknown and evolutionary" (Hendry, 1995, p.xxix). Furthermore, in "the forecasting context ... the degree of data congruence or non-congruence of a model is neither necessary nor sufficient for forecasting success or failure." (Clements and Hendry, 1999, p.xxiv). These authors develop auto-regressive forecasting models which use difference data, and non-causal variables, showing that such models can out-perform historically calibrated and causally-relevant models.

In macro-economic modelling

> there is genuine uncertainty about how good a model is, even within the sample. Moreover, since the economy is evolving, we can take it for granted that the data generation process will change in the forecast period, causing any model of it to become mis-specified over that period, and this is eventually the main problem in economic forecasting. (Bardsen *et al,* 2005, p.246)

Bardsen *et al* (2005) build on the work of Clements and Hendry (1999), showing the advantage of difference models in forecasting the behavior of non-stationary systems.

The current paper studies the situation where contextual understanding indicates that fundamental structural change may occur in the future, but these changes are not yet manifested in the data. Adaptive or learning strategies will not reveal future structural breaks. For instance, in section 5 we consider forecasting European Central Bank interest rates after the 9/11 Al Qaida terror attacks but before the Central Bank has revised its lending policy. We use info-gap models for uncertainty in the future transition matrix of linear time-varying systems. Info-gap models are non-probabilistic and particularly suited to modelling major structural change for which historical data provide no evidence.

We address these questions: Why, and in what sense, is fidelity to data an insufficient criterion for forecast success? What characteristics of forecasting models are indicative of forecast success? How can the forecaster use contextual information which suggests that the system properties will change fundamentally but in ways which are not yet manifested in historical data?

Two foci of uncertainty are present in forecasting with uncertain models: noisy data as well as fundamental errors in model structure. The classical paradigm of optimality — maximize the fidelity of the model to the data by minimizing an error function — is not directly applicable to this situation because this paradigm addresses only one focus of uncertainty: data noise. In this paper we show how both foci of uncertainty can be managed.

A basic theorem of info-gap theory (Ben-Haim, 2006) asserts the irrevocable trade-off between enhancing fidelity of a model to data, and ameliorating the structural errors in the model itself. Robustness to model error decreases as the analyst demands greater fidelity to the data; maximal fidelity entails minimal robustness to model mis-specification. Treating both data-noise and model-error requires a compromise.

What this means is that forecasts cannot, realistically, be as good as the data themselves suggest, when models are wrong in unknown ways. Thus the key insight of Gauss and Legendre when they invented least-squares estimation — let the data themselves dictate the fidelity to the model — is inappropriate when the model structures which underlie the forecast are uncertain. The implication is that fidelity to data should be satisficed rather than optimized when formulating a forecasting model. Satisficed (sub-optimal) fidelity rarely entails a unique forecasting algorithm, so there remains an additional degree of freedom in the forecasting process which can be devoted to maximizing the robustness to model uncertainty. Our first theorem asserts the existence of sub-optimal models which are better forecasters than optimal-fidelity models. Our second theorem asserts that the info-gap robustness function (which is calculable with very limited information) is a good proxy for the probability of forecast success (which cannot be calculated without knowning probability distributions). Info-gap theory thus provides a tool for selecting good models for forecasting, and for understanding why models that optimally fit historical data may be poor models for forecasting.

Our general results are developed in section 2, illustrated with a 1-dimension example in section 3, and explored for an N-dimensional system in section 4. A simple financial example is discussed in section 5, and our results are summarized in section 6. All proofs appear in section 7.

2 Forecasting with Info-Gap Uncertainty

2.1 The Estimated System and its Uncertainty

Consider a system whose state is characterized by $y_t \in \Re^N$ and whose average behavior evolves as:

$$y_t = A_t y_{t-1} \tag{1}$$

Zero-mean, additive, random disturbances are ignored and all other inputs are incorporated in the multi-dimensional state vector y_t.

The solution of this system is:

$$y_{T+k} = \left(\prod_{i=1}^{k} A_{T+i} \right) y_T \tag{2}$$

where the product operator is defined as lefthand matrix multiplication: $\prod_{i=1}^{k} A_{T+i} = A_{T+k} \prod_{i=1}^{k-1} A_{T+i}$.

Historical data indicate that the coefficient matrix A_t is constant for $t \le T$. The best-estimate of this constant matrix is $\widetilde{A}$. It is anticipated that A_t will remain constant, though auxiliary information and understanding — economic reasoning, engineering judgment, medical intuition, etc. — suggest that its elements could vary systematically. The uncertainty in the future values of the elements of A_t, for $t > T$, is represented by an **info-gap model** (Ben-Haim, 2006), $\mathcal{U}(\alpha, \widetilde{A})$, $\alpha \ge 0$, which is an unbounded family of nested sets of coefficient matrices A. Info-gap models obey two axioms:

$$\text{Nesting:} \quad \alpha < \alpha' \quad \text{implies} \quad \mathcal{U}(\alpha, \widetilde{A}) \subset \mathcal{U}(\alpha', \widetilde{A}) \tag{3}$$

$$\text{Contraction:} \quad \mathcal{U}(0, \widetilde{A}) = \{\widetilde{A}\} \tag{4}$$

'Nesting' implies that the uncertainty set $\mathcal{U}(\alpha, \widetilde{A})$ becomes more inclusive as α increases, which endows α with its meaning of an horizon of uncertainty. 'Contraction' implies that, in the absence of uncertainty ($\alpha = 0$), only the estimated matrix $\widetilde{A}$ occurs. The horizon of uncertainty is unknown so α can take any non-negative value.

A common info-gap model, among many possibilities, is the fractional-error model:

$$\mathcal{U}(\alpha, \widetilde{A}) = \left\{ A_t, t > T : \ \widetilde{A}_{ij} - \alpha v_{ij} \le [A_t]_{ij} \le \widetilde{A}_{ij} + \alpha w_{ij}, \ i, j = 1, \ldots, N \right\}, \quad \alpha \ge 0 \tag{5}$$

where v_{ij} and w_{ij} are known non-negative "uncertainty weights" which might be chosen from boundaries of confidence intervals. If our anticipation of future stability of A_t is correct, so that $\alpha = 0$, then A_t remains at the estimated value $\widetilde{A}$. However, as the horizon of uncertainty, α, increases, the coefficients at each time step can vary in possibly asymmetric intervals defined in eq.(5). These intervals are of unknown size since the horizon of uncertainty, α, is unbounded. There are many other types of info-gap models of uncertainty, all obeying the axioms of nesting and contraction.

2.2 Forecasting with Slope Adjustment

The analyst is concerned that, while the estimate of A_t in eq.(1) is historically stable, contextual and other knowledge suggest that A_t might tend to vary systematically in the future, as expressed by an info-gap model such as eq.(5). We do not know, however, by how much the coefficients may vary, or how that variation would emerge over time. Despite this uncertainty, we need to forecast the future values of some of the state variables.

As compensation for potential future variation in the coefficients one might consider the following "slope-adjusted" predictor:

$$y_t^{\rm s} = B y_{t-1}^{\rm s} \tag{6}$$

where B is a constant matrix which we are free to choose. The question is: how to choose B?

The vector of average forecast errors of this forecast model for time $t = T + k$ (ignoring zero-mean, additive, random disturbances), based knowledge of y_T, is:

$$\eta_k(B, A_t) = y_{T+k}^{\rm s} - y_{T+k} = \left(B^k - \prod_{i=1}^{k} A_{T+i} \right) y_T \tag{7}$$

Judicious choice of B can reliably compensate for deviation of A_{T+i} from $\widetilde{A}$, as we will show. Specifically, we will prove two theorems. Theorem 1 asserts the existence of sub-optimal forecast models, such as eq.(6), which are more robust to future uncertainty in the system, than the optimally-estimated model $\widetilde{A}$. Theorem 2 specifies very general conditions in which the probability of forecast success increases with increasing robustness of the forecast model. Combining these theorems shows when forecast will be more reliable (against future surprises) with sub-optimal than with optimally-estimated forecast models.

2.3 Definition of the Robustness Function

We require that the absolute error of the average forecast of the mth state variable, k time steps after the last measurement, not exceed $\varepsilon_{\rm c}$:

$$|\eta_{k,m}(B, A_t)| \le \varepsilon_{\rm c} \tag{8}$$

That is, ignoring zero-mean, additive, random disturbances, we wish to satisfice the average forecast error of $y_{T+k,m}$ at the value $\varepsilon_{\rm c}$. We are willing to accept some bias in the forecast, but no more than $\varepsilon_{\rm c}$. Likewise, we are willing to accept fluctuation in the forecast error (due to fluctuation in the coefficient matrices), but again no more than $\varepsilon_{\rm c}$.

The **info-gap robustness** of a forecast with coefficient matrix B in eq.(6) is the greatest horizon of uncertainty α up to which all realizations of the actual coefficients A_{T+i} in an info-gap model $\mathcal{U}(\alpha, \widetilde{A})$ satisfice the forecast error at $\varepsilon_{\rm c}$:

$$\widehat{\alpha}(B, \varepsilon_{\rm c}) = \max \left\{ \alpha : \left(\max_{\substack{A_{T+i} \in \mathcal{U}(\alpha, \widetilde{A}) \\ i=1, \ldots, k}} |\eta_{k,m}(B, A_t)| \right) \le \varepsilon_{\rm c} \right\} \tag{9}$$

We define $\widehat{\alpha}(B, \varepsilon_{\rm c}) = 0$ if the set of α-values in eq.(9) is empty.

2.4 1-Step and k-Step Forecast

From eq.(2) we see that a k-step process can be viewed as a 1-step process with coefficient matrix $A^{(k)} = \prod_{i=1}^{k} A_{T+i}$. If the individual matrices A_{T+i} belong to an info-gap model, then the product matrix $A^{(k)}$ also belongs to an info-gap model, as asserted in the following lemma. All proofs for this section appear in section 7.1.

Lemma 1 *Let $\mathcal{U}(\alpha, \widetilde{A})$ be an info-gap model for square matrices A, that is, the sets $\mathcal{U}(\alpha, \widetilde{A})$ obey the nesting and contraction axioms, eqs.(3) and (4). Define the following family of sets of matrices:*

$$\mathcal{U}_k(\alpha, \widetilde{A}^k) = \left\{ A = \prod_{i=1}^{k} A_i : \ A_i \in \mathcal{U}(\alpha, \widetilde{A}) \right\}, \quad \alpha \ge 0 \tag{10}$$

$\mathcal{U}_k(\alpha, \widetilde{A}^k)$ *obeys the nesting and contraction axioms and is therefore an info-gap model.*

What this lemma implies is that the k-step evolution of a system with info-gap-uncertain matrices can be treated as a 1-step system with a different info-gap model for the uncertain matrices.[1] In particular, any conclusion about the 1-step forecast robustness which depends only upon the generic properties of info-gap models — nesting and contraction — holds for k-step forecasts for any value of k. In other words, we need only consider 1-step forecasts as long as we consider completely generic info-gap models for uncertainty in the coefficient matrices. On the other hand, a result which depends for instance on the explicit interval-bound structure of the info-gap model in eq.(5), would not necessarily hold for k-step forecasts, since the info-gap model in eq.(10) does not possess interval-bound structure if it is derived from the 1-step matrices in eq.(5).

2.5 Evaluating the Robustness Function

In this section we derive the 1-step robustness function for a generic info-gap model. The results of this section apply also to the robustness of k-step forecast if B, $\widetilde{A}$ and $\mathcal{U}(\alpha, \widetilde{A})$ are replaced by B^k, $\widetilde{A}^k$ and $\mathcal{U}_k(\alpha, \widetilde{A}^k)$, as explained in section 2.4.

The robustness in eq.(9) can be written:

$$\widehat{\alpha}(B, \varepsilon_{\rm c}) = \max \left\{ \alpha : \left(\max_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \eta_{1,m}(B, A_{T+1}) \right) \le \varepsilon_{\rm c} \right. \\ \left. \text{and} \left(\min_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \eta_{1,m}(B, A_{T+1}) \right) \ge -\varepsilon_{\rm c} \right\} \tag{11}$$

The 1-step forecast error for the mth state variable, from eq.(7), is:

$$\eta_{1,m}(B, A_{T+1}) = \underbrace{\sum_{n=1}^{N} [B - \widetilde{A}]_{mn} y_{T,n}}_{\delta} - \sum_{n=1}^{N} [A_{T+1} - \widetilde{A}]_{mn} y_{T,n} \tag{12}$$

which defines the quantity δ, which can be either positive or negative and is controlled by the analyst through the choice of the forecast matrix B.

The concepts of 'coherence' and 'anti-coherence' will arise subsequently, in definition 2 and the discussion preceding it. We now introduce two functions which will be used extensively. Define:

$$\theta_{\rm c}(\alpha) = \max_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^{N} [A_{T+1} - \widetilde{A}]_{mn} y_{T,n} \tag{13}$$

$$\theta_{\rm a}(\alpha) = - \min_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^{N} [A_{T+1} - \widetilde{A}]_{mn} y_{T,n} \tag{14}$$

A large value for $\theta_{\rm c}(\alpha)$ implies a 'coherence' between the fluctuations of A_{T+1} and the elements of y_T. Likewise, a large value for $\theta_{\rm a}(\alpha)$ implies an 'anti-coherence' between these entities. Note that the contraction axiom implies that $\theta_{\rm a}(0) = \theta_{\rm c}(0) = 0$. The nesting axiom then implies that $\theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha)$ are non-negative and monotonic for all $\alpha \ge 0$.

Now, from eqs.(12)–(14), the robustness for 1-step forecast of the mth state variable, eq.(11), is:

$$\widehat{\alpha}(B, \varepsilon_{\rm c}) = \max \{ \alpha : \ \delta + \theta_{\rm a}(\alpha) \le \varepsilon_{\rm c} \quad \text{and} \quad -\delta + \theta_{\rm c}(\alpha) \le \varepsilon_{\rm c} \} \tag{15}$$

This can be used to conveniently evaluate the robustness function, once the functions $\theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha)$ are known. Define:

$$\varepsilon(\alpha) = \max \{ \delta + \theta_{\rm a}(\alpha), \ -\delta + \theta_{\rm c}(\alpha) \} \tag{16}$$

A plot of α vs. $\varepsilon(\alpha)$ is the same as a plot of $\widehat{\alpha}(B, \varepsilon_{\rm c})$ vs. $\varepsilon_{\rm c}$. This is illustrated schematically in fig. 1. The vertical axis is α or $\widehat{\alpha}(B, \varepsilon_{\rm c})$, while the horizontal axis is $\varepsilon(\alpha)$ or $\varepsilon_{\rm c}$.

[1] The complexity of the info-gap models $\mathcal{U}_k(\alpha, \widetilde{A}^k)$ increases greatly with increasing k. This can be exploited for artistic purposes as illustrated in (Ben-Haim, 1997).

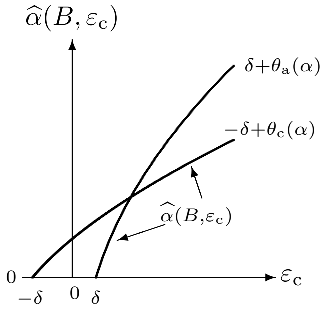

Figure 1: Robustness function based on eqs.(15) and (16).

2.6 Crossing of Robustness Curves and the Advantage of Sub-Optimal Models

If the analyst is correct in the anticipation that A_t will be constant at $\widetilde{A}$ in the future, then the k-step prediction error for the mth state variable is:

$$\eta_{k,m}(B,\widetilde{A}) = \underbrace{\sum_{n=1}^{N} \left(B^k - \widetilde{A}^k\right)_{mn} y_{T,n}}_{\varepsilon^\star} \tag{17}$$

Since we are free to choose the matrix B, one is tempted to choose $B = \widetilde{A}$ in order to minimize the anticipated prediction error $\varepsilon^\star$.

However, recall the definition of δ in eq.(12), for the k-step case so that B and $\widetilde{A}$ are replaced by B^k and $\widetilde{A}^k$ as explained at the beginning of section 2.5. We note that $\delta = \varepsilon^\star$. The robustness (to uncertainty in A_t) vanishes for prediction error ε_c equal to or less than δ for any choice of B, as implied by eq.(15) and illustrated in fig. 1 (unless $\theta_a(\alpha) = \theta_c(\alpha) = 0$ for $\alpha > 0$). This means that prediction error as low as δ, for any choice of B, cannot be reliably attained. We must "migrate up" the robustness curve, to larger prediction error, in order to achieve positive robustness against the unknown future values of the coefficient matrix A_t.

We need a definition before we can state our first main result.

Definition 1 *Given a non-negative ε_c and the functions $\theta_a(\alpha)$ and $\theta_c(\alpha)$ defined in eqs.(13) and (14), define $\delta_\times(\varepsilon_c)$ and $\alpha_\times(\varepsilon_c)$ from:*

$$\theta_a(\alpha_\times) + \delta_\times = \theta_c(\alpha_\times) - \delta_\times \tag{18}$$

$$\theta_a(\alpha_\times) + \delta_\times = \varepsilon_c \tag{19}$$

$\delta_\times(\varepsilon_c)$ is the value of δ at which the intersection of $\theta_a(\alpha) + \delta$ with $\theta_c(\alpha) - \delta$ occurs at $\varepsilon_c = \theta_a(\alpha) + \delta = \theta_c(\alpha) - \delta$.

The forecasting model based on $\widetilde{A}$ is optimal in the sense of having maximal fidelity to historical data, and being maximally reliable if the future system is the same as the past system. However, we are studying situations in which the future behavior of the system may deviate from the past. The following theorem asserts that any positive forecast error can be achieved with greater robustness against uncertainty in A_t with particular sub-optimal forecasters, $B \neq \widetilde{A}$, than with the historically optimal forecaster $B = \widetilde{A}$.

Theorem 1 *There exist sub-optimal models for 1-step forecasting which are more robust than optimal models.*
Given: *y_T is not identically zero. $\mathcal{U}(\alpha,\widetilde{A})$ is an info-gap model for uncertainty in the coefficient matrix in eq.(1). $\theta_c(\alpha)$ and $\theta_a(\alpha)$, defined in eqs.(13) and (14), are continuous, at least one is unbounded, and either $\theta_c(\alpha) \geq \theta_a(\alpha)$ or $\theta_a(\alpha) \geq \theta_c(\alpha)$ for all $\alpha > 0$.*

Then: *for any $\varepsilon_c > 0$ for which $\delta_\times(\varepsilon_c) \neq 0$, there is a $B \neq \widetilde{A}$ such that:*

$$\widehat{\alpha}(B,\varepsilon_c) > \widehat{\alpha}(\widetilde{A},\varepsilon_c) \tag{20}$$

where these are robustness functions for 1-step forecast.

We will consider examples in sections 3 and 4.

Lemma 1 implies that theorem 1 holds for the robustness functions of k-step forecasts. If $\mathcal{U}(\alpha, \widetilde{A})$ is an info-gap model for uncertainty in the coefficient matrix in eq.(1), then define $\mathcal{U}_k(\alpha, \widetilde{A}^k)$ as in eq.(10), which is an info-gap model according to lemma 1. Now define $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$ in eqs.(13) and (14) with respect to $\mathcal{U}_k(\alpha, \widetilde{A}^k)$. k-step forecasts are now 1-step forecasts with a more complicated info-gap model, and theorem 1 applies.

2.7 Robustness and Probability of Forecast Success

We have demonstrated, in theorem 1, that there exist forecasting models which are more robust to model-uncertainty than the historically optimal forecasting model. That is, there exist sub-optimal models whose forecast error is within a specified interval for a wider range of future transition matrices, than the optimal forecast model. This, in itself, is a definite advantage of these sub-optimal forecasters. However, this greater robustness does not necessarily imply that the *probability* that the forecast error is within a specified interval is greater for the sub-optimal forecaster. However, 'greater robustness' *does* imply 'greater probability' if there is sufficient coherence between the info-gap model of uncertainty (which is non-probabilistic but based on fragmentary contextual understanding), and the probability distribution of the model error (which is unknown). We develop the concept of coherence in this section, and present a theorem which establishes sufficient conditions for greater robustness to imply greater probability of successful forecast.

We first derive an expression for the probability of 1-step forecast success, and then connect it to the robustness of the forecasting model in eq.(6). Our results are applicable to k-step forecasting if B, $\widetilde{A}$ and $\mathcal{U}(\alpha, \widetilde{A})$ are replaced by B^k, $\widetilde{A}^k$ and $\mathcal{U}_k(\alpha, \widetilde{A}^k)$, as explained in section 2.4.

The 1-step forecast of the mth state variable is successful if the forecast errs no more than $\varepsilon_{\rm c}$, which is eq.(8) with $k = 1$. From eq.(7) this forecast error is:

$$\eta_{1,m}(B, A_{T+1}) = \sum_{n=1}^{N} [B - A_{T+1}]_{mn}\, y_{T,n} \tag{21}$$

The condition for forecast success, $|\eta_{1,m}| \le \varepsilon_{\rm c}$, can be written explicitly as:

$$-\varepsilon_{\rm c} + \sum_{n=1}^{N} [B - \widetilde{A}]_{mn} y_{T,n} \le \underbrace{\sum_{n=1}^{N} \left[A_{T+1} - \widetilde{A}\right]_{mn} y_{T,n}}_{u} \le \varepsilon_{\rm c} + \sum_{n=1}^{N} [B - \widetilde{A}]_{mn} y_{T,n} \tag{22}$$

which defines the variable u. Recalling the definition of δ in eq.(12), the condition for forecast success in eq.(22) becomes:

$$\delta - \varepsilon_{\rm c} \le u \le \delta + \varepsilon_{\rm c} \tag{23}$$

$\widetilde{A}$ and y_T are known when the 1-step forecast is made, but A_{T+1} and u are unknown. Let $F(\cdot)$ denote the unknown cumulative probability distribution (cpd) of u, with probability density function (pdf) $f(\cdot)$. Eq.(23) implies that the probability of 1-step forecast success, with forecasting model based on B, is:

$$P_{\rm s}(B) = F(\delta + \varepsilon_{\rm c}) - F(\delta - \varepsilon_{\rm c}) \tag{24}$$

Differentiating this we obtain:

$$\frac{{\rm d}P_{\rm s}(B)}{{\rm d}\delta} > 0 \quad \text{if and only if} \quad f(\delta + \varepsilon_{\rm c}) > f(\delta - \varepsilon_{\rm c}) \tag{25}$$

δ can be changed by altering the matrix B of the forecasting model. Roughly speaking, eq.(25) asserts that the probability of 1-step forecast success is increased by increasing δ if and only if $f(u)$ is skewed to values above δ as compared to values below δ. Likewise, the probability of forecast success increases by decreasing δ if and only if $f(u)$ is skewed to the left around δ.

Recall the definitions of $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$ in eqs.(13) and (14). These functions are known before the forecast model is chosen and the forecast is made. $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$ depend on the info-gap model of uncertainty, $\mathcal{U}(\alpha, \widetilde{A})$. The info-gap model is chosen by the analyst to reflect contextual understanding about how the transition matrix might vary in the future. For instance, in section 5 we will consider uncertainty in an interest rate whose future value is unknown, but for which economic understanding suggests the direction in which it might vary. If this contextual understanding is sufficiently realistic, then the info-gap model reveals something about the future fluctuations. Specifically, the functions $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$ reveal something about the random fluctuation of the variable u, as we now explain.

If $\theta_{\rm c}(\alpha) > \theta_{\rm a}(\alpha)$ for all $\alpha > 0$, then the info-gap model $\mathcal{U}(\alpha, \widetilde{A})$ implies that fluctuations of A_{T+1} around $\widetilde{A}$ will tend to be coherent with the signs of the elements of the last observed state vector, y_T. If the info-gap model is sufficiently realistic, then this implies that the random variable u, defined in eq.(22), will tend to be positive rather than negative. In other words, the probability density, $f(u)$, will tend to be increasing around $u = 0$.

If $\theta_{\rm c}(\alpha) < \theta_{\rm a}(\alpha)$ for all $\alpha > 0$, then fluctuations of A_{T+1} around $\widetilde{A}$ will tend to be anti-coherent with the signs of the elements of the last observed state vector, y_T. This implies that $f(u)$ will tend to be decreasing around $u = 0$.

In other words, *if the info-gap model is sufficiently realistic,* then the functions $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$ reveal something about the probability distribution of the random variable u. This motivates the following definition.

Definition 2 *The info-gap model* $\mathcal{U}(\alpha, \widetilde{A})$ *and the cpd* $F(u)$ *are* **coherent** *at* $(\delta, \varepsilon_{\rm c})$ *given state vector* y_T *if:*

$$[\theta_{\rm c}(\alpha) - \theta_{\rm a}(\alpha)]\,[f(\delta + \varepsilon_{\rm c}) - f(\delta - \varepsilon_{\rm c})] \geq 0 \quad \textit{for all } \alpha > 0 \tag{26}$$

with strict inequality for at least one value of α.

The info-gap model $\mathcal{U}(\alpha, \widetilde{A})$ *and the cpd* $F(u)$ *are* **anti-coherent** *at* $(\delta, \varepsilon_{\rm c})$ *given state vector* y_T *if:*

$$[\theta_{\rm c}(\alpha) - \theta_{\rm a}(\alpha)]\,[f(\delta + \varepsilon_{\rm c}) - f(\delta - \varepsilon_{\rm c})] \leq 0 \quad \textit{for all } \alpha > 0 \tag{27}$$

with strict inequality for at least one value of α.

From our discussion of the relations $\theta_{\rm c}(\alpha) > \theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha) < \theta_{\rm a}(\alpha)$ just before definition 2, we see that we should expect realistic info-gap models to be coherent with the cpd of u for values of δ near the origin.

We now state the following theorem, which asserts that, if $\mathcal{U}(\alpha, \widetilde{A})$ and $F(u)$ are coherent, then the probability of success for 1-step forecasts is increased by increasing the info-gap robustness. As explained at the beginning of this section, this theorem also holds for k-step forecasts with the appropriate notational changes.

Theorem 2 *The probability of 1-step forecast success increases with increasing robustness of the forecast model, for coherent info-gap models. Robustness and forecast success are inversely related for anti-coherent info-gap models.*

Given: *y_T is not identically zero, $\varepsilon_{\rm c}$ is non-negative, $\delta_\times(\varepsilon_{\rm c})$ is non-zero, and the functions $\theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha)$ are continuous and at least one of them is unbounded.*

If $\mathcal{U}(\alpha, \widetilde{A})$ *and* $F(u)$ *are* coherent *at* $\varepsilon_{\rm c}$ *for some* $|\delta| < |\delta_\times(\varepsilon_{\rm c})|$*, and* $\widehat{\alpha}(B, \varepsilon_{\rm c}) > 0$ *for this* δ *and* $\varepsilon_{\rm c}$*,* **then:**

$$\frac{{\rm d}\widehat{\alpha}(B, \varepsilon_{\rm c})}{{\rm d}\delta} > 0 \quad \textit{if and only if} \quad \frac{{\rm d}P_{\rm s}(B)}{{\rm d}\delta} > 0 \tag{28}$$

If $\mathcal{U}(\alpha, \widetilde{A})$ *and* $F(u)$ *are* anti-coherent *at* $\varepsilon_{\rm c}$ *for some* $|\delta| < |\delta_\times(\varepsilon_{\rm c})|$*, and* $\widehat{\alpha}(B, \varepsilon_{\rm c}) > 0$ *for this* δ *and* $\varepsilon_{\rm c}$*,* **then:**

$$\frac{{\rm d}\widehat{\alpha}(B, \varepsilon_{\rm c})}{{\rm d}\delta} > 0 \quad \textit{if and only if} \quad \frac{{\rm d}P_{\rm s}(B)}{{\rm d}\delta} < 0 \tag{29}$$

3 Example: 1-Dimensional System

The system. Consider a scalar system whose average behavior evolves as:

$$y_t = \lambda_t y_{t-1} \tag{30}$$

where the historical data indicate that the growth coefficient λ_t is constant for $t \le T$. The best-estimate of this constant coefficient is $\widetilde{\lambda}$, which is positive. It is anticipated that λ will remain constant, though auxiliary information and understanding suggest that it could drift upwards. The uncertainty in the future values of λ_t, for $t > T$, is represented by a fractional-error info-gap model:

$$\mathcal{U}(\alpha, \widetilde{\lambda}) = \left\{ \lambda_t, t > T : \ 0 \le \frac{\lambda_t - \widetilde{\lambda}}{\widetilde{\lambda}} \le \alpha \right\}, \quad \alpha \ge 0 \tag{31}$$

This means that, if our anticipation is correct so that $\alpha = 0$, then λ_t remains at the value $\widetilde{\lambda}$. However, as the horizon of uncertainty, α, increases, the coefficient at each time step can vary in the interval $[\widetilde{\lambda}, \ (1+\alpha)\widetilde{\lambda}]$. The value of α is unknown. Eq.(31) is a special case of eq.(5)

The slope-adjusted forecaster. As a rough compensation for potential future increase in the growth coefficient one might consider the following slope-adjusted predictor, where $\ell \ge \widetilde{\lambda}$:

$$y_t^{\mathrm{s}} = \ell y_{t-1}^{\mathrm{s}} \tag{32}$$

The robustness of k-step forecast with growth coefficient ℓ, defined in eq.(9) and derived in section 7.2, is:

$$\widehat{\alpha}(\ell, \varepsilon_{\mathrm{c}}) = \begin{cases} 0 & \text{if } \varepsilon_{\mathrm{c}} \le (\ell^k - \widetilde{\lambda}^k) y_T \\ \left(\dfrac{\varepsilon_{\mathrm{c}} + \ell^k y_T}{\widetilde{\lambda}^k y_T} \right)^{1/k} - 1 & \text{else} \end{cases} \tag{33}$$

Crossing of robustness curves. For this example one can readily show that:

$$\theta_{\mathrm{a}}(\alpha) = \min[\widetilde{\lambda}, \ \alpha\widetilde{\lambda}] y_T \tag{34}$$

$$\theta_{\mathrm{c}}(\alpha) = \alpha\widetilde{\lambda} y_T \tag{35}$$

Hence, if $y_T \ne 0$, the conditions of theorem 1 are satisfied. Thus sub-optimal coefficients, ℓ, can be found whose robustness curves cross the robustness curve for the model with the estimated coefficient, $\widetilde{\lambda}$. We will see examples.

Robustness and probability of success. If $y_T \ne 0$ and if the info-gap model is coherent with $F(u)$, then the conditions of theorem 2 are satisfied and the robustness is equivalent to the probability of forecast success.

Furthermore, if $y_T > 0$ and if λ_{T+k} in fact cannot be less than $\widetilde{\lambda}$, then the info-gap model is necessarily coherent with $F(u)$, as we now explain. From eqs.(34) and (35) we conclude that $\theta_{\mathrm{c}}(\alpha) \ge \theta_{\mathrm{a}}(\alpha)$. Furthermore, note that $\delta = (\ell^k - \widetilde{\lambda}^k) y_T$ for k-step forecast. Thus eq.(33) shows that positive robustness implies that $\delta - \varepsilon_{\mathrm{c}} < 0$. The random variable u is non-negative because $\lambda_{T+k} \ge \widetilde{\lambda}$, so $f(\delta - \varepsilon_{\mathrm{c}}) = 0$. Hence if $f(u) > 0$ for all $u \ge 0$, then $f(\delta + \varepsilon_{\mathrm{c}}) > f(\delta - \varepsilon_{\mathrm{c}})$. Thus the info-gap model is necessarily coherent with $F(u)$.

Numerical example. Fig. 2 shows numerical evaluation of the robustness function for the slope-adjusted forecasting model, eq.(33). The left frame is for 1-step prediction, $k = 1$, the middle frame predicts 2 steps ahead, $k = 2$, and the left frame predicts 3 steps, $k = 3$. The lowest curve in each frame is the robustness of the nominal forecaster, $\ell = \widetilde{\lambda} = 1.05$. The slope-adjusted growth parameter ℓ increases by 0.05 with each higher curve. The horizontal axis is the satisficed forecast error, ε_{c}, normalized to the nominal forecast value, $\widetilde{\lambda}^k y_T$.

In all three frames we observe the curve-crossing phenomenon. The nominal optimal predictor has greater robustness than the sub-optimal predictors at low error (small ε_{c}), while the sub-optimal predictors have greater robustness than the nominal predictor at higher error. The analyst must choose a suitable predictor by considering the trade-off between robustness and forecast-error. In other words, the analyst must choose between a model with higher robustness at larger ε_{c}, or a model with nonzero robustness at smaller ε_{c}.

In the 1-step forecast shown in the left frame, the slope-adjusted predictors are far more robust than the nominal predictor for all levels of forecast error ε_{c} except for very small values. For instance, consider 5% fractional forecast error, $\varepsilon_{\mathrm{c}}/\widetilde{\lambda}^k y_T = 0.05$. For $k = 1$ we find that $\widehat{\alpha}(1.05, \varepsilon_{\mathrm{c}}) = 0.050$ (bottom curve), and $\widehat{\alpha}(1.2, \varepsilon_{\mathrm{c}}) = 0.19$ (top curve). That is, at 5% forecast error, the slope-adjusted forecaster with $\ell = 1.2$ is robust to 19% error in actual realization of the growth coefficient

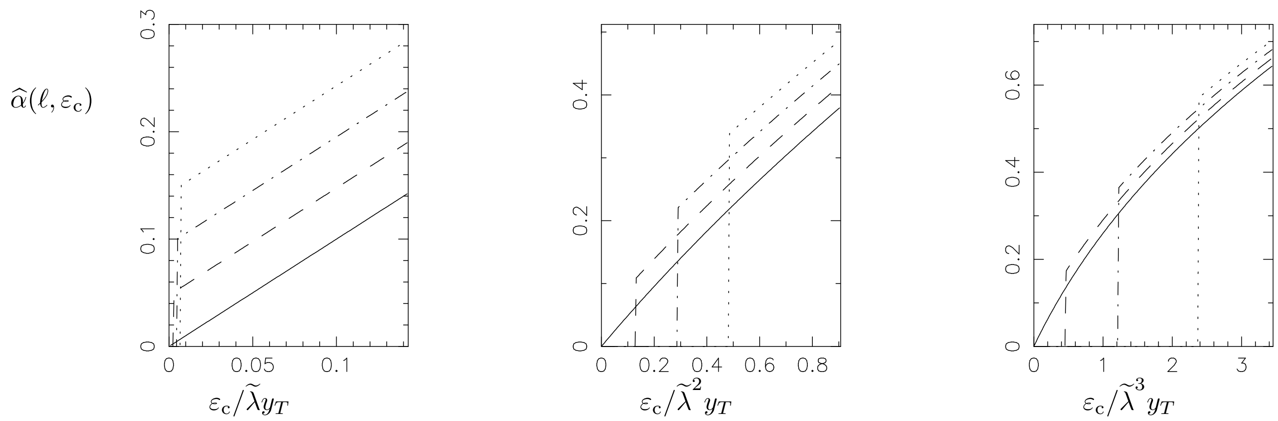

Figure 2: Robustness vs. normalized forecast error, eq.(33), for $\ell = 1.05$, 1.1, 1.15, 1.2 from bottom to top curve. $\widetilde{\lambda} = 1.05$, $y_T = 1$. $k = 1$ (left), 2(mid), 3(right).

λ_{T+1}, while the nominal forecaster is robust to only 5% error. The slope-adjusted predictor is about 4 times more robust than the nominal predictor.

The middle and right frames show that the robustness premium of the slope-adjusted forecaster, $\ell > \widetilde{\lambda}$, compared to the nominal predictor, $\ell = \widetilde{\lambda}$, becomes smaller as the horizon of the prediction increases. For instance, consider the 2-step forecast at normalized prediction error $\varepsilon_{\rm c}/\widetilde{\lambda}^k y_T = 0.5$. The robustnesses are $\widehat{\alpha}(1.2, \varepsilon_{\rm c}) = 0.34$ compared against $\widehat{\alpha}(1.05, \varepsilon_{\rm c}) = 0.22$. That is, the slope-adjusted predictor is 50% more robust, but only at 50% prediction error.

The robustness premium of the slope-adjusted predictor is even lower at $k = 3$. With normalized prediction error $\varepsilon_{\rm c}/\widetilde{\lambda}^k y_T = 2.4$, we have robustnesses $\widehat{\alpha}(1.2, \varepsilon_{\rm c}) = 0.57$ compared against $\widehat{\alpha}(1.05, \varepsilon_{\rm c}) = 0.50$. The slope-adjusted predictor is only 14% more robust, and only at 240% prediction error.

4 Multi-Dimensional System

System and uncertainty model. The average system behavior is described by eq.(1), and the coefficient matrix A_t is estimated up to time T to be $\widetilde{A}$. The uncertainty in future values of A_t is represented by the fractional-error info-gap model of eq.(5).

Crossing of robustness curves. For the info-gap model of eq.(5) one can readily derive $\theta_{\rm c}(\alpha)$ and $\theta_{\rm a}(\alpha)$, defined in eqs.(13) and (14). Let $s_n = 1$ if $y_{T,n} \geq 0$ and $s_n = 0$ otherwise. Then:

$$\theta_{\rm c}(\alpha) = \alpha \underbrace{\sum_{n=1}^{N} [v_{mn}(1 - s_n) + w_{mn} s_n] \, |y_{T,n}|}_{\tau_{\rm c}} \tag{36}$$

$$\theta_{\rm a}(\alpha) = \alpha \underbrace{\sum_{n=1}^{N} [w_{mn}(1 - s_n) + v_{mn} s_n] \, |y_{T,n}|}_{\tau_{\rm a}} \tag{37}$$

$\tau_{\rm c}$ and $\tau_{\rm a}$ are non-negative and determined by the info-gap model and the known value of the last observed state vector y_T. We know their values when the forecast is made, but we cannot influence them. $\tau_{\rm c}$ expresses a "coherency" between the info-gap uncertainty about A_{T+1} and the state vector y_T: the sum in eq.(36) matches upper-envelope weights, w_{mn}, with positive elements of y_T, and lower-envelope weights, v_{mn}, with negative elements of y_T. $\tau_{\rm a}$ expresses an "anti-coherency": w_{mn}'s are matched with negative elements of y_T; v_{mn}'s with positive elements.

The conditions of theorem 1 are satisfied if $y_T \neq 0$, so sub-optimal forecast models, eq.(6), exist whose robustness curves cross the robustness curve of the estimated model. We will demonstrate this explicitly.

Robustness and probability of forecast success. The info-gap model $\mathcal{U}(\alpha, \widetilde{A})$ is coherent with the cpd $F(u)$ at $(\delta, \varepsilon_{\rm c})$, according to definition 2, if $(\tau_{\rm c} - \tau_{\rm a})[f(\delta + \varepsilon_{\rm c}) - f(\delta - \varepsilon_{\rm c})] > 0$. The further conditions of theorem 2 are guaranteed so that, if coherency holds, then the probability of forecast success increases with increasing robustness of the forecast model.

For $\theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha)$ in eqs.(36) and (37), and from lemma 2 in section 7.1, we find:

$$\delta_{\times}(\varepsilon_{\rm c}) = \frac{\tau_{\rm c} - \tau_{\rm a}}{\tau_{\rm c} + \tau_{\rm a}} \varepsilon_{\rm c} \tag{38}$$

So, if $\tau_{\rm c} \neq \tau_{\rm a}$ and if coherency holds, then the probability of forecast success increases with increasing robustness for $|\delta| < |\delta_{\times}|$.

Robustness function for 1-step forecast.

We now state explicit expressions for the robustness function for 1-step-forecasts. These results are derived in section 7.3. Let $\text{sgn}(x)$ denote the algebraic sign of x.

δ, defined in eq.(12), depends on the matrix B, it can be positive or negative, and we can choose its value since we are free to choose B. The sign of δ expresses a coherency between the last observation, y_T, and the difference between the forecasting and estimated models, B and $\widetilde{A}$. δ is positive if B_{mn} exceeds $\widetilde{A}_{mn}$ for positive $y_{T,n}$ and if B_{mn} is less than $\widetilde{A}_{mn}$ for negative $y_{T,n}$. δ will be negative if the reverse correlations dominate.

Define:

$$\tau_2 = \max\{\tau_{\rm c},\ \tau_{\rm a}\}, \quad \tau_1 = \min\{\tau_{\rm c},\ \tau_{\rm a}\}, \quad \varepsilon_{\times} = \frac{\tau_2 + \tau_1}{\tau_2 - \tau_1}|\delta|, \quad \widehat{\alpha}_{\times} = \frac{2|\delta|}{\tau_2 - \tau_1} \tag{39}$$

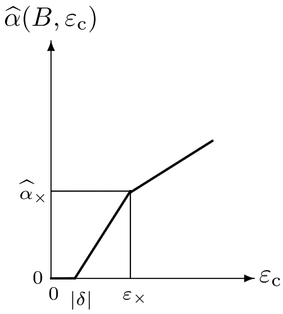

Figure 3: Robustness function of eq.(40).

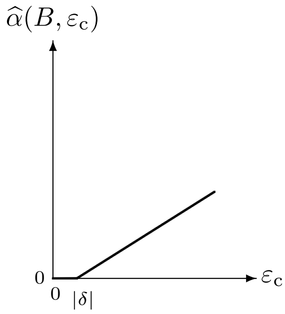

Figure 4: Robustness function of eq.(41).

We can now state explicit expressions for the robustness function.

If $\text{sgn}(\delta) = \text{sgn}(\tau_{\rm c} - \tau_{\rm a})$ then (fig. 3):

$$\widehat{\alpha}(B, \varepsilon_{\rm c}) = \begin{cases} 0 & \varepsilon_{\rm c} \leq |\delta| \\ \dfrac{\varepsilon_{\rm c} - |\delta|}{\tau_1} & |\delta| \leq \varepsilon_{\rm c} < \varepsilon_{\times} \\ \dfrac{\varepsilon_{\rm c} + |\delta|}{\tau_2} & \varepsilon_{\times} \leq \varepsilon_{\rm c} \end{cases} \tag{40}$$

The robustness at the "kink" is $\widehat{\alpha}_{\times}$, defined in eq.(39) and shown in fig. 3.

If $\text{sgn}(\delta) \neq \text{sgn}(\tau_{\rm c} - \tau_{\rm a})$ then (fig. 4):

$$\widehat{\alpha}(B, \varepsilon_{\rm c}) = \begin{cases} 0 & \varepsilon_{\rm c} \leq |\delta| \\ \dfrac{\varepsilon_{\rm c} - |\delta|}{\tau_2} & |\delta| \leq \varepsilon_{\rm c} \end{cases} \tag{41}$$

The two solutions converge to one another when $\tau_{\rm a} = \tau_{\rm c}$ since the "kink" at $(\varepsilon_{\times}, \widehat{\alpha}_{\times})$ runs out to infinity.

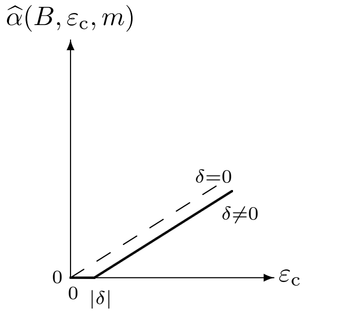

Figure 5: Robustness function if $\mathrm{sgn}(\delta) \neq \mathrm{sgn}(\tau_c - \tau_a)$, eq.(41).

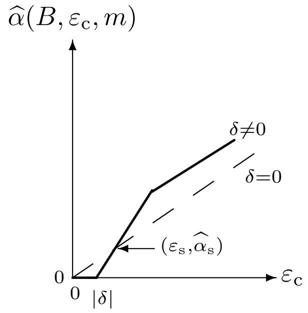

Figure 6: Robustness function if $\mathrm{sgn}(\delta) = \mathrm{sgn}(\tau_c - \tau_a)$, eq.(40).

The advantage of sub-optimal models. Robustness curves are shown again in figs. 5 and 6 for zero and non-zero values of δ. If B is such that $\mathrm{sgn}(\delta) \neq \mathrm{sgn}(\tau_c - \tau_a)$, fig. 5, then the robustness with $\delta = 0$ (dashed line) exceeds the robustness with B (solid line). Nonetheless, the robustness vanishes at the anticipated prediction error, which is $\varepsilon^\star = 0$ for $B = \widetilde{A}$. However, we will shortly see that $\mathrm{sgn}(\delta) \neq \mathrm{sgn}(\tau_c - \tau_a)$ can always be avoided and that greater robustness can always be obtained.

When B is such that $\mathrm{sgn}(\delta) = \mathrm{sgn}(\tau_c - \tau_a)$, fig. 6, then the robustness curve for $\delta \neq 0$ (thick lines) crosses the robustness curve for $\delta = 0$ (dashed line). Let $(\varepsilon_s, \widehat{\alpha}_s)$ denote the point at which these robustness curves cross. The choice $B = \widetilde{A}$ will be more robust to future uncertainty in A_t than the choice of B whose non-zero δ-value is shown in the figure, for prediction-error aspirations $\varepsilon_c \leq \varepsilon_s$. However, the robustness will also be no greater than $\widehat{\alpha}_s$ which may be quite small, depending on the value of δ. If prediction error as small as ε_s is highly desired and if robustness no greater than $\widehat{\alpha}_s$ is adequate, then $B = \widetilde{A}$ is the predictor of choice. However, a predictor based on $B \neq \widetilde{A}$ has greater robustness than $B = \widetilde{A}$ for any prediction-error $\varepsilon_c > \varepsilon_s$, and its robustness will exceed $\widehat{\alpha}_s$. Thus, if large robustness is needed and if prediction error in excess of ε_s (which may be quite small) can be tolerated, then a sub-optimal forecast model, $B \neq \widetilde{A}$, will be preferred.

Finally, let us note that δ is freely chosen, while τ_a and τ_c are observed but not controlled. Comparison of eqs.(40) and (41) shows that, for any $\varepsilon_c > |\delta|$, greater robustness is always obtained by choosing $\mathrm{sgn}(\delta) = \mathrm{sgn}(\tau_c - \tau_a)$, eq.(40), rather than $\mathrm{sgn}(\delta) \neq \mathrm{sgn}(\tau_c - \tau_a)$, eq.(41). Furthermore, for any $\varepsilon_c > 0$, greater robustness is obtained with $\delta \neq 0$ which has been chosen so that $\varepsilon_s < \varepsilon_c$, where ε_s is shown in fig. 6. In short, a sub-optimal forecasting model, $B \neq \widetilde{A}$, can always be chosen with greater robustness than the estimated model, $B = \widetilde{A}$. This is the gist of theorem 1.

Date	Interest rate	Implied λ
1 Jan 1999	4.50	
9 Apr 1999	3.50	0.778
5 Nov 1999	4.00	1.143
4 Feb 2000	4.25	1.063
17 Mar 2000	4.50	1.059
28 Apr 2000	4.75	1.056
9 Jun 2000	5.25	1.105
28 Jun 2000	5.25	1.000
1 Sep 2000	5.50	1.048
6 Oct 2000	5.75	1.045
11 May 2001	5.50	0.957
31 Aug 2001	5.25	0.955

Table 1: Interest rates for overnight loans at the European Central Bank (marginal lending facility). Source: http://www.ecb.int/stats/monetary/rates/html/index.en.html

5 Example: European Central Bank Interest Rates

Table 1 shows historical values of interest rates for overnight loans (marginal lending facility) from the European Central Bank (ECB) to commercial banks for the period from 1999 (when the ECB started making euro loans) until August 2001. Typically, the ECB changes its rate by 25 basis points (100 basis points = 1 percent) though changes as large as 50 or 100 basis points are observed. Furthermore, the 14 months from June 2000 to August 2001 show more than usual stability, with a mean rate of 5.4% and a standard deviation of only 0.19%.

Now consider the forecasting problem on September 12th, 2001, one day after the al-Qaida attack in the US and before the next interest-rate announcement. Economic reasoning suggests that interest rates will go down in order to counteract anticipated economic contraction resulting from the terror attacks. But by how much? How should the forecaster choose, and justify, the forecast?

We will describe the evolution of the interest rate with the 1-dimensional system in eq.(30). The estimated transition coefficient based on the observed behavior of the ECB from June 2000 to August 2001 is $\widetilde{\lambda} = 1$. Our forecast will be based on the slope-adjusted model of eq.(32). It is likely that the ECB will reduce the interest rate, so that λ_T is expected to be less than $\widetilde{\lambda}$ and the forecast should use a value of ℓ in eq.(32) less than $\widetilde{\lambda}$. Use the info-gap model of eq.(5) for uncertainty in λ_T with $w = 0$ and $v = \widetilde{\lambda}$:

$$\mathcal{U}(\alpha, \widetilde{\lambda}) = \left\{ \lambda_T : \ (1-\alpha)\widetilde{\lambda} \le \lambda_T \le \widetilde{\lambda} \right\}, \quad \alpha \ge 0 \tag{42}$$

The horizon of uncertainty, α, is the unknown fractional error in $\widetilde{\lambda}$ as an estimate of the transition coefficient for the next announced interest rate.

The last observed interest rate, $y_T = 5.25$, is positive so we find, from eqs.(36) and (37), $\tau_{\mathrm{a}} = \widetilde{\lambda} y_T$ and $\tau_{\mathrm{c}} = 0$. From eq.(12) we find $\delta = (\ell - \widetilde{\lambda}) y_T < 0$. From eq.(39) we have $\varepsilon_\times = (\widetilde{\lambda} - \ell) y_T$ and $\widehat{\alpha}_\times = 2(\widetilde{\lambda} - \ell)/\widetilde{\lambda}$. Eq.(40) applies (note that $|\delta| = \varepsilon_\times$) so the robustness is:

$$\widehat{\alpha}(\ell, \varepsilon_{\mathrm{c}}) = \begin{cases} 0 & \varepsilon_{\mathrm{c}} < \varepsilon_\times \\ \dfrac{\varepsilon_{\mathrm{c}}}{\widetilde{\lambda} y_T} + \dfrac{\widetilde{\lambda} - \ell}{\widetilde{\lambda}} & \varepsilon_\times \le \varepsilon_{\mathrm{c}} \end{cases} \tag{43}$$

This robustness function shows the same dis-continuity and curve-crossing behavior observed in fig. 2. Namely, $\ell < \widetilde{\lambda}$ is more robust than $\widetilde{\lambda}$ for positive estimation error.

How should the analyst choose, and explain, the sub-optimal forecast coefficient ℓ? The choice will depend on both the robustness-to-uncertainty which the analyst requires, $\widehat{\alpha}_{\mathrm{d}}$, and the acceptable estimation error ε_{c}.

Table 1 shows values of λ implied by each interest rate change, evaluated as the ratio of new to previous interest rate.[2] The mean and standard deviation of the implied λ over all the observations are 1.047 and 0.093, while the mean and standard deviation over the past five observations are 1.001 and 0.041. Robustness to 10% variation in λ would seem fairly safe, so one might adopt $\widehat{\alpha}_{\mathrm{d}} = 0.1$.

In order to forecast the next announced interest rate, we invert eq.(43) to find ℓ:

$$\ell = \left(1 + \frac{\varepsilon_{\mathrm{c}}}{\widetilde{\lambda} y_T} - \widehat{\alpha}_{\mathrm{d}} \right) \widetilde{\lambda} \tag{44}$$

Requiring fractional forecast error $\varepsilon_{\mathrm{c}}/\widetilde{\lambda} y_T = 0.02$ and adopting $\widehat{\alpha}_{\mathrm{d}} = 0.1$, we choose $\ell = 0.92$ resulting in a forecasted interest rate $y^{\mathrm{s}}_{T+1} = \ell y_T = 4.83$. This is a reasonable interest rate, which an informed observer of the ECB might select based only on contemplation of the out-of-data evidence (the 9/11 al-Qaida attacks). However, the info-gap robust-satisficing forecast guarantees that the fractional forecast error will be no greater than 0.02 for errors up to 10% in the historical estimate of λ. More generally, the analyst chooses the forecasting model (ℓ) in light of the performance which is required of the forecast (ε_{c} and $\widehat{\alpha}_{\mathrm{d}}$).

[2] The announcement intervals are not uniform. This will not impact our analysis which will be based on the last six announcements which are very nearly constant.

In fact the ECB announced an interest rate of 4.75 on September 18th, 2001. Thus the fractional forecast error was $|4.83 - 4.75|/5.25 = 0.015$. An informed observer of the ECB may well have predicted this value without any quantitative analysis. The purpose of this simple example is to illustrate the procedure for constructing an info-gap model and using a robustness function to make a forecast with specified robustness to uncertainty and specified prediction-error.

6 Summary

We have used info-gap decision theory to implement a robust-satisficing approach to forecasting the future behavior of a system whose trajectory evolves in unknown ways over time, and about which some limited out-of-data contextual knowledge is available. We consider linear discrete-time systems, and use info-gap models to represent uncertainty in the future time-varying transition matrix. Our theorems are true for general info-gap models, requiring only the properties of nesting and contraction. The forecaster requires the average forecast of a specific state variable to be within a specified interval around the correct value.

Our first theorem asserts that models with sub-optimal fidelity to historical data can, as forecasters, be more robust to future uncertainty in the system model than models with optimal fidelity to data. More specifically, for any positive forecast error, $\varepsilon_{\rm c}$, there exist sub-optimal models which can guarantee forecast error no greater than $\varepsilon_{\rm c}$ for a larger range of future transition matrices than the historically optimal model. The theorem indicates how such models can be constructed. We have not explored the important question of finding sub-optimal models with maximum robustness. Nor have we explicitly considered error in the dimension of the model, but only error in the future values of the time-varying transition matrix. However, uncertain dimensionality can be analyzed with our method by allowing uncertain variation of rows and columns which are strictly zero in the estimated transition matrix. We have not considered uncertain non-linearities.

Our second theorem identifies conditions in which the probability of forecast success increases with increasing robustness to model error. The theorem is based on the idea of "coherency" between the info-gap model of uncertainty in the transition matrix, A_t, and the unknown probability density, $f(u)$, of a particular scalar function of A_t. The info-gap model is constructed to represent an infinity of possible future trends in A_t. Coherency holds if the sign of $\theta_{\rm c}(\alpha) - \theta_{\rm a}(\alpha)$, which depends on the info-gap model but not on $f(u)$, corresponds to a particular skewness of $f(u)$. That is, the theorem asserts that, if the info-gap model captures a particular weak feature of the unknown pdf of A_t, then info-gap robustness is positively correlated with the probability of forecast success.

Combining these two theorems we see that, in well defined circumstances, we can identify models which have sub-optimal fidelity to historical data and yet have greater probability of forecast success than optimal-fidelity models.

7 Proofs

7.1 For Section 2

Proof of lemma 1. (1) $\mathcal{U}(\alpha, \widetilde{A})$ obeys the nesting axiom so $\mathcal{U}_k(\alpha, \widetilde{A}^k)$ becomes more inclusive as α increases. Hence $\mathcal{U}_k(\alpha, \widetilde{A}^k)$ obeys nesting. (2) $\mathcal{U}(\alpha, \widetilde{A})$ obeys the contraction axiom so $\mathcal{U}(0, \widetilde{A}) = \{\widetilde{A}\}$. Hence $\mathcal{U}_k(0, \widetilde{A}^k) = \{\widetilde{A}^k\}$ so $\mathcal{U}_k(\alpha, \widetilde{A}^k)$ obeys contraction. ∎

Lemma 2 Given: *a non-negative $\varepsilon_{\rm c}$ and at least one of the functions $\theta_{\rm a}(\alpha)$ and $\theta_{\rm c}(\alpha)$ is continuous and unbounded.*

Then *$\delta_\times(\varepsilon_{\rm c})$ and $\alpha_\times(\varepsilon_{\rm c})$ are finite and satisfy:*

$$\varepsilon_{\rm c} = \frac{\theta_{\rm a}(\alpha_\times) + \theta_{\rm c}(\alpha_\times)}{2} \qquad (45)$$

$$\delta_\times = \frac{\theta_{\rm c}(\alpha_\times) - \theta_{\rm a}(\alpha_\times)}{2} \qquad (46)$$

Proof of lemma 2. The assumption about unboundedness and the nesting and contraction axioms imply that $\theta_{\rm a}(0) = \theta_{\rm c}(0) = 0$ and at least one of them increases with increasing α. Since

$\varepsilon_c \geq 0$ and either $\theta_a(\alpha)$ or $\theta_c(\alpha)$ is continuous and unbounded, there is a non-negative $\alpha_\times$ which satisfies eq.(45). With this $\alpha_\times$, define $\delta_\times$ from eq.(46). It is readily shown that this $\alpha_\times$ and $\delta_\times$ satisfy eqs.(18) and (19). ∎

Proof of theorem 1. We prove the theorem for the case that:

$$\theta_a(\alpha) \geq \theta_c(\alpha) \tag{47}$$

The proof for the reverse inequality is essentially the same.[3]

The nesting axiom implies that $\theta_a(0) = \theta_c(0) = 0$. Since $\varepsilon_c > 0$ and $\theta_a(\alpha)$ and $\theta_c(\alpha)$ are continuous and at least one is unbounded there is a positive α which satisfies:

$$\frac{\theta_a(\alpha) + \theta_c(\alpha)}{2} = \varepsilon_c \tag{48}$$

Call the largest such value $\alpha_\times$. Now define:

$$\delta = \frac{\theta_c(\alpha_\times) - \theta_a(\alpha_\times)}{2} \tag{49}$$

which, by eq.(47), is not positive. From lemma 2 we see that δ in eq.(49) equals $\delta_\times(\varepsilon_c)$ which, by supposition, is not zero. Thus $\delta < 0$. Also: $-\delta < \varepsilon_c$. From the definition of δ in eq.(12), and since $y_T \neq 0$, we see that one can construct a matrix B, different from $\widetilde{A}$, for which the value of δ is given by eq.(49).

For δ in eq.(49) we find:

$$\delta + \theta_a(\alpha_\times) = \varepsilon_c = -\delta + \theta_c(\alpha_\times) \tag{50}$$

Thus from eq.(15) and (50) we find that $\widehat{\alpha}(B, \varepsilon_c)$ is the solution for $\alpha_\times$ of:

$$\delta + \theta_a(\alpha_\times) = \varepsilon_c \tag{51}$$

Recall that $\delta = 0$ when $B = \widetilde{A}$. Thus, from eqs.(15) and (47) we find that $\widehat{\alpha}(\widetilde{A}, \varepsilon_c)$ is the solution for α of:

$$\theta_a(\alpha) = \varepsilon_c \tag{52}$$

Since $\varepsilon_c - \delta > \varepsilon_c$ and $\theta_a(\alpha)$ is continuous and unbounded we conclude that there exist solutions to eqs.(51) and (52), and that the solution of eq.(51) exceeds the solution of eq.(52). This proves eq.(20). ∎

Lemma 3 Given: *At least one of the functions $\theta_a(\alpha)$ and $\theta_c(\alpha)$ is continuous and unbounded.* **Then:**

$$\theta_c(\alpha) \geq \theta_a(\alpha) \text{ for all } \alpha > 0 \quad \text{if and only if} \quad \delta_\times(\varepsilon_c) \geq 0 \text{ for all } \varepsilon_c \geq 0 \tag{53}$$

Proof of lemma 3. The lemma results directly from eq.(46). ∎

Lemma 4 Given: *y_T is not identically zero, ε_c is non-negative, and the functions $\theta_a(\alpha)$ and $\theta_c(\alpha)$ are continuous and at least one of them is unbounded.* **Then,** *for any δ for which $\widehat{\alpha}(B, \varepsilon_c)$ is positive:*

$$\frac{d\widehat{\alpha}(B, \varepsilon_c)}{d\delta} < 0 \quad \text{if} \quad \delta > \delta_\times(\varepsilon_c) \tag{54}$$

$$\frac{d\widehat{\alpha}(B, \varepsilon_c)}{d\delta} > 0 \quad \text{if} \quad \delta < \delta_\times(\varepsilon_c) \tag{55}$$

Proof of lemma 4. (1) Examination of the definition of δ in eq.(12) shows that, since $y_T \neq 0$, one can construct a matrix B to obtain any desired value of δ.

From the nesting axiom we see that $\theta_a(\alpha)$ and $\theta_c(\alpha)$ cannot decrease as α increases. That is, $\theta_a(\alpha)$ and $\theta_c(\alpha)$ have non-negative slopes.

From lemma 2, $\delta_\times(\varepsilon_c)$ and $\alpha_\times(\varepsilon_c)$ exist. From eqs.(18) and (19) recall that:

$$\theta_a(\alpha_\times) + \delta_\times = \varepsilon_c = \theta_c(\alpha_\times) - \delta_\times \tag{56}$$

[3] The differences are minor: δ, defined in eq.(49), is positive and less than ε_c. Then, use $-\delta + \theta_c(\alpha_\times) = \varepsilon_c$ instead of eq.(51). Finally, use $\varepsilon_c + \delta > \varepsilon_c$ following eq.(52).

(2) Suppose $\delta < \delta_\times(\varepsilon_c)$. Suppose there exists an α_1 value for which:

$$\theta_a(\alpha_1) + \delta = \varepsilon_c \tag{57}$$

Since $\theta_a(\alpha)$ has non-negative slope, we see from eq.(56) any such α_1 must satisfy:

$$\alpha_1 > \alpha_\times(\varepsilon_c) \tag{58}$$

Likewise, from eq.(56) and since $\theta_c(\alpha)$ is continuous with non-negative slope, if there exists an α_2 value for which:

$$\theta_c(\alpha_2) - \delta = \varepsilon_c \tag{59}$$

then:

$$\alpha_2 < \alpha_\times(\varepsilon_c) \tag{60}$$

Now recall the formulation of the robustness function in eq.(15), and that $\theta_a(\alpha)$ and $\theta_c(\alpha)$ are continuous. Eqs.(58) and (60) imply that $\alpha_2 < \alpha_1$ so that eq.(59) is the binding constraint and the robustness is the greatest α satisfying:

$$\theta_c(\alpha) - \delta \le \varepsilon_c \tag{61}$$

From this it results that the robustness increases as δ increases because $\theta_c(\alpha)$ has non-negative slope, which proves eq.(55).

(3) Suppose $\delta > \delta_\times(\varepsilon_c)$. Suppose there is an α_2 satisfying:

$$\theta_c(\alpha_2) - \delta = \varepsilon_c \tag{62}$$

Since $\theta_c(\alpha)$ has non-negative slope, we see from eqs.(56) and (62) that any such α_2 must satisfy:

$$\alpha_2 > \alpha_\times(\varepsilon_c) \tag{63}$$

Likewise, from eq.(56) and because $\theta_a(\alpha)$ is continuous with non-negative slope, if there exists a value of α_1 such that:

$$\theta_a(\alpha_1) + \delta = \varepsilon_c \tag{64}$$

then:

$$\alpha_1 < \alpha_\times(\varepsilon_c) \tag{65}$$

Recall the robustness function in eq.(15), and that $\theta_a(\alpha)$ and $\theta_c(\alpha)$ are continuous. Eqs.(63) and (65) imply that $\alpha_1 < \alpha_2$ so that eq.(64) is the binding constraint and the robustness is the greatest α satisfying:

$$\theta_a(\alpha) + \delta \le \varepsilon_c \tag{66}$$

From this it results that the robustness decreases as δ increases because $\theta_a(\alpha)$ is continuous with non-negative slope, which proves eq.(54). ∎

Proof of theorem 2.

1. We first consider eq.(28). From eq.(25) and the assumed coherency of $\mathcal{U}(\alpha, \widetilde{A})$ with $F(u)$ for this δ we have:

$$\theta_c(\alpha) \ge \theta_a(\alpha) \text{ for all } \alpha > 0 \quad \text{if and only if} \quad \frac{\mathrm{d}P_s(B)}{\mathrm{d}\delta} > 0 \tag{67}$$

with strict inequality for some α.

To complete the proof we must show that, for this δ:

$$\frac{\mathrm{d}\widehat{\alpha}(B, \varepsilon_c)}{\mathrm{d}\delta} > 0 \tag{68}$$

if and only if:

$$\theta_c(\alpha) \ge \theta_a(\alpha) \text{ for all } \alpha > 0 \tag{69}$$

with strict inequality for some α.

1.1 Suppose that eq.(69) holds. Thus lemma 2 implies that $\delta_\times(\varepsilon_c) \ge 0$. Hence $|\delta| < |\delta_\times(\varepsilon_c)|$ implies that $\delta < \delta_\times(\varepsilon_c)$. Now eq.(55) of lemma 4 implies eq.(68).

1.2 Suppose that eq.(68) holds. Hence lemma 4 implies that the δ-value in eq.(68) satisfies:

$$\delta \le \delta_\times(\varepsilon_{\rm c}) \tag{70}$$

Thus, since $|\delta| < |\delta_\times(\varepsilon_{\rm c})|$, we conclude that $\delta_\times(\varepsilon_{\rm c}) > 0$. From lemma 2 we conclude that $\theta_{\rm c}(\alpha_\times) - \theta_{\rm a}(\alpha_\times) > 0$. Hence, since the sign of $\theta_{\rm c}(\alpha) - \theta_{\rm a}(\alpha)$ is constant for all $\alpha > 0$, we conclude that eq.(69) holds.

2. We now consider eq.(29). From eq.(25) and the assumed anti-coherency of $\mathcal{U}(\alpha, \widetilde{A})$ with $F(u)$ for this δ we have:

$$\theta_{\rm c}(\alpha) \ge \theta_{\rm a}(\alpha) \text{ for all } \alpha > 0 \quad \text{if and only if} \quad \frac{{\rm d}P_{\rm s}(B)}{{\rm d}\delta} < 0 \tag{71}$$

with strict inequality for some α.

To complete the proof we must again show that eqs.(68) and (69) are equivalent. The argument in step 1.2 is still valid, which completes the proof of the theorem. ∎

7.2 For Section 3: Derivation of Eq.(33)

Let $M(\alpha)$ denote the inner maximum in the definition of the robustness, eq.(9). The info-gap model $\mathcal{U}(\alpha, \widetilde{A})$ becomes more inclusive as α increases, thus $M(\alpha)$ increases with increasing α. The robustness is the greatest value of α at which $M(\alpha) \le \varepsilon_{\rm c}$. Thus the robustness is the greatest α which satisfies $M(\alpha) = \varepsilon_{\rm c}$. In other words, at any fixed ℓ, if $M(\alpha)$ is strictly increasing in α, then $M(\alpha)$ is the inverse of the robustness function $\widehat{\alpha}(\ell, \varepsilon_{\rm c})$:

$$M(\alpha) = \varepsilon_{\rm c} \quad \text{implies} \quad \widehat{\alpha}(\ell, \varepsilon_{\rm c}) = \alpha \tag{72}$$

At horizon of uncertainty α, each coefficient λ_{T+i} of the info-gap model in eq.(31) varies in the interval $\widetilde{\lambda} \le \lambda_{T+i} \le (1+\alpha)\widetilde{\lambda}$. The absolute maximum forecast error at horizon of uncertainty α, $M(\alpha)$, is obtained at one of the extremes: either $\lambda_{T+i} = \widetilde{\lambda}$ or $\lambda_{T+i} = (1+\alpha)\widetilde{\lambda}$ for all $i = 1, \ldots, k$, whichever is greater:

$$M(\alpha) = \max\left\{ (\ell^k - \widetilde{\lambda}^k) y_T, \ [(1+\alpha)^k \widetilde{\lambda}^k - \ell^k] y_T \right\} \tag{73}$$

From this we find:

$$M(\alpha) = \left\{ \begin{array}{ll} (\ell^k - \widetilde{\lambda}^k) y_T & \text{if } 0 \le \alpha < \left(\frac{2\ell^k}{\widetilde{\lambda}^k} - 1 \right)^{1/k} - 1 \\ [(1+\alpha)^k \widetilde{\lambda}^k - \ell^k] y_T & \text{if } \left(\frac{2\ell^k}{\widetilde{\lambda}^k} - 1 \right)^{1/k} - 1 < \alpha \end{array} \right. \tag{74}$$

Inverting $M(\alpha)$ we obtain eq.(33).

7.3 For Section 4: Derivation of Eqs.(40) and (41)

We now derive the robustness function for 1-step forecasts, eqs.(40) and (41).

The 1-step prediction error for the mth state variable is given in eq.(21). The 1-step robustness can be written:

$$\begin{aligned} \widehat{\alpha}(B, \varepsilon_{\rm c}) = \max \Bigg\{ \alpha : \quad & \max_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^N [A_{T+1}]_{mn} \, y_{T,n} \le \varepsilon_{\rm c} + \sum_{n=1}^N B_{mn} y_{T,n} \\ \text{and} \quad & \min_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^N [A_{T+1}]_{mn} \, y_{T,n} \ge -\varepsilon_{\rm c} + \sum_{n=1}^N B_{mn} y_{T,n} \Bigg\} \end{aligned} \tag{75}$$

At horizon of uncertainty α in the info-gap model of eq.(5), the inner maximum in eq.(75) is obtained with the following choice of $[A_{T+1}]_{mn}$:

$$[A_{T+1}]_{mn} \quad = \quad \left\{ \begin{array}{ll} \widetilde{A}_{mn} - \alpha v_{mn} & \text{if } y_{T+1,n} < 0 \\ \widetilde{A}_{mn} + \alpha w_{mn} & \text{if } y_{T+1,n} \ge 0 \end{array} \right. \tag{76}$$

As before define $s_n = 1$ if $y_{T,n} \geq 0$ and $s_n = 0$ otherwise. The inner maximum in eq.(75) becomes:

$$\max_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^{N} [A_{T+1}]_{mn} \, y_{T,n} = \sum_{n=1}^{N} \widetilde{A}_{mn} y_{T,n} + \alpha \underbrace{\sum_{n=1}^{N} [v_{mn}(1-s_n) + w_{mn} s_n] \, |y_{T,n}|}_{\tau_{\rm c}} \tag{77}$$

where $\tau_{\rm c}$ is the "coherency" between the last measurement, y_T, and the info-gaps in A_{T+1}, defined in eq.(36).

Similarly, at horizon of uncertainty α, the inner minimum in eq.(75) is obtained with the following choice of $[A_{T+1}]_{mn}$:

$$[A_{T+1}]_{mn} \quad = \quad \left\{ \begin{array}{ll} \widetilde{A}_{mn} - \alpha v_{mn} & \text{if } y_{T+1,n} \geq 0 \\ \widetilde{A}_{mn} + \alpha w_{mn} & \text{if } y_{T+1,n} < 0 \end{array} \right. \tag{78}$$

The inner minimum in eq.(75) becomes:

$$\min_{A_{T+1} \in \mathcal{U}(\alpha, \widetilde{A})} \sum_{n=1}^{N} [A_{T+1}]_{mn} \, y_{T,n} = \sum_{n=1}^{N} \widetilde{A}_{mn} y_{T,n} - \alpha \underbrace{\sum_{n=1}^{N} [w_{mn}(1-s_n) + v_{mn} s_n] \, |y_{T,n}|}_{\tau_{\rm a}} \tag{79}$$

where $\tau_{\rm a}$ is the "anti-coherency" between y_T and the info-gaps in A_{T+1}, defined in eq.(37).

Recalling the definition of δ in eq.(12), eqs.(75), (77) and (79) can be combined as:

$$\widehat{\alpha}(B, \varepsilon_{\rm c}, m) = \max \left\{ \alpha : \ \alpha \leq \frac{\varepsilon_{\rm c} + \delta}{\tau_{\rm c}} \ \text{ and } \ \alpha \leq \frac{\varepsilon_{\rm c} - \delta}{\tau_{\rm a}} \right\} \tag{80}$$

We are now ready to explain eq.(40) and then eq.(41). We must consider 4 cases.

$\tau_{\rm c} > \tau_{\rm a}$ and $\delta \geq 0$, shown in fig. 7. The thin lines show the two linear constraints which α must obey in eq.(80). The heavy line is the intersection of these constraints and thus defines the robustness, which is eq.(40) for this choice of δ.

$\tau_{\rm c} < \tau_{\rm a}$ and $\delta \leq 0$, shown in fig. 8. Once again the thin lines show the two linear constraints which α must obey in eq.(80). The heavy line is the intersection of these constraints and thus defines the robustness, which is again eq.(40) for this choice of δ.

$\tau_{\rm c} > \tau_{\rm a}$ and $\delta \leq 0$, shown in fig. 9. The thin lines show the two linear constraints which α must obey in eq.(80). The right-most line is the intersection of these constraints and thus defines the robustness, which is eq.(41) for this choice of δ.

$\tau_{\rm c} < \tau_{\rm a}$ and $\delta \geq 0$, shown in fig. 10. Again the thin lines show the two linear constraints which α must obey in eq.(80). The right-most line is the intersection of these constraints and thus defines the robustness, which is again eq.(41) for this choice of δ.

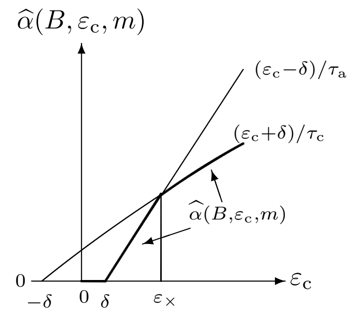

Figure 7: Robustness function for $\tau_{\rm c} > \tau_{\rm a}$ and $\delta \geq 0$.

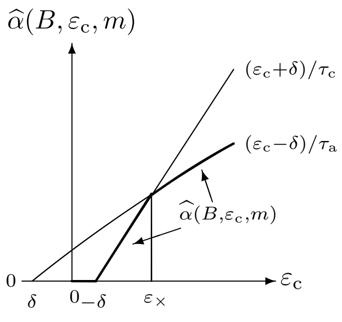

Figure 8: Robustness function for $\tau_{\rm c} < \tau_{\rm a}$ and $\delta \leq 0$.

Acknowledgements

The author is pleased to acknowledge valuable comments and suggestions by Lior Davidovitch and Miriam Zacksenhouse.

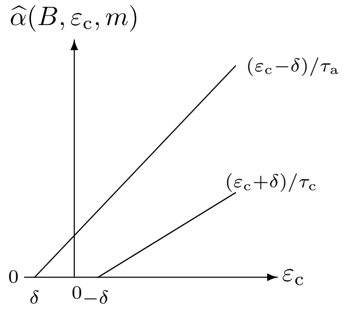

Figure 9: Robustness function for $\tau_c > \tau_a$ and $\delta \leq 0$.

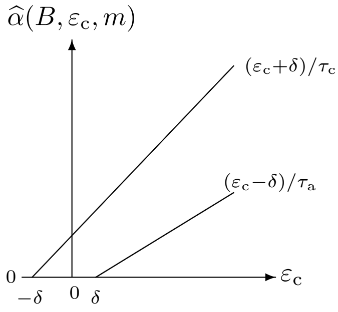

Figure 10: Robustness function for $\tau_c < \tau_a$ and $\delta \geq 0$.

8 References

Bardsen, G, Eitrheim, Ø, Jansen, ES and Nymoen, R, 2005, *The Econometrics of Macroeconomic Modelling,* Oxford University Press.

Ben-Haim, Y, 1997, Art and the Exponential: Graphical Exploration of the Matrix Exponential Function, *Leonardo,* Journal of the International Society for Arts, Sciences and Technology, 30: 49–52.

Ben-Haim, Y, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty,* 2nd ed., Academic Press, London.

Clements, MP, and Hendry, DF, 1999, *Forecasting Non-Stationary Economic Time Series,* MIT Press.

Hendry, DF, 1995, *Dynamic Econometrics,* Oxford University Press.

Parker, EN, 2006, Shielding space travelers, *Scientific American,* March, pp.41–47.

Smith, DJ, 2006, Predictability and preparedness in influenza control, *Science,* vol. 312, pp.392–394.