UNIVERSITÀ DEGLI STUDI DI BERGAMO

Department of Economics

WORKING PAPERS

Can we estimate macroforecasters' mis-behavior?

Emilio Zanetti Chini

January 2023 - WP N. 20 Year 2023

Working papers – Department of Economics
n. 20

Can we estimate macroforecasters' mis-behavior?

Emilio Zanetti Chini

Università degli Studi di Bergamo
2023

Can we estimate macroforecasters' mis-behavior? / Emilio Zanetti Chini – Bergamo: Università degli Studi di Bergamo, 2023.
Working papers of Department of Economics, n. 20
ISSN: 2974-5586
DOI: 10.13122/WPEconomics_20

Progetto grafico: Servizi Editoriali – Università degli Studi di Bergamo
Università degli Studi di Bergamo
via Salvecchio, 19
24129 Bergamo
Cod. Fiscale 80004350163
P. IVA 01612800167

https://aisberg.unibg.it/handle/10446/235330

Can we estimate macroforecasters' mis-behavior?

Emilio Zanetti Chini*

University of Bergamo
Department of Economics
Via dei Caniana, 2 - 24127, Bergamo (ITALY)
e-mail: emilio.zanettichini@unibg.it

December 2022

Abstract

We answer positively to this question by using Maximum Lq-Likelihood (or Deformed Likelihood) estimator. This is based on a parameter which measures the aggregate quote of judgment in the forecasting (game-based) system formed by three players—Forecaster, Policy Maker and Reality. For the first time in econometric literature, we apply this estimator to a dynamic system and derive a robust version of the Kalman Filter—the Deformed Kalman Filter (DKF). The evidence from U.S. data suggests that the judgmental dynamics exists and is correlated (but not coincident) with the phases of the Business Cycle. Furthermore its knowledge improves in-sample as well as out-of-sample estimation.

Keywords: Deformed Likelihood, Dynamic Systems, Judgment, Repeated Games, Robust Filtering.
JEL: C1, C2, C5, E3, E7.

*The Author is particularly grateful to Giovanni Urga and Salvatore Piccolo for their support. He also thanks Barbara Annicchiarico, Marcelle Chauvet, Raffaella Giacomini, Ana Maria Herrera, Giuseppe Ragusa, Francesco Ravazzolo and Tom Stark for their feedback, as all the participants to the 2nd Ventotene Workshop in Macoreconomics, the 41st International Symposium on Forecasting 2021, the internal Seminar Series in Mathematics and Statistics of the University of Bergamo, the IWEEE2022 in Rimini and the IAAE2022 in London. In some of that occasion the paper had been presented with the title "*Measuring Unobserved Strategic Judgment*". Finally, the Author is in debt with doctors, nurses and workers of Policlinico "S. Matteo" of Pavia, without whose uncompensated cares this paper would not have been written.

"*It's easier to invent than to observe.*" (Gioacchino Rossini).

"*If you torture data long enough, it will confess to anything.*" (Ronald H. Coase).

1 Introduction

The difficulty in observing and correctly interpreting a phenomenon is a non-trivial aspect of any scientific knowledge. When the Italian composer formulated his statement, these capabilities—and their understanding—were a privilege for a restricted élite. Thus, despite its humor, the statement should be considered as a (partial) truth. The growing pervasiveness of information technology and the rapid acceleration and development of statistical tools for data processing since the mid 20th Century made the incentive for exploiting the capability of data science superior to the incentive to think of new approaches to Economics, for the first time in Social Science history. Thus the Nobel Laureate's statement might be considered less paroxysmal than suggested by a superficial interpretation.

The bias in the survey of the U.S. economy's professional forecasters is an example of mismatching among observation and interpretation in contemporary Macroeconomics. Figure 1 illustrates the case of 1-quarter-ahead forecasts of the U.S. Real GDP (RGDP, henceforth): namely, panel (a) considers the bias across time—the spread among realizations and Greenbook estimates[1]—while panel (b) investigates the same bias in a cross-sectional perspective—that is, the spread among the realizations and the 1-quarter ahead estimates for individuals corresponding to the 75th and 25th percentiles of the distribution of survey respondents (denoted RGDP-P75 and RGDP-P25, respectively). The spread among data and Greenbook forecasts is pervasive both in time and in cross-sectional dimension, even if considering the difference among the 75th and 25th percentile' survey respondents as a centered measure to make our investigation robust to extremely optimistic/pessimistic forecasters.

The existence of the experts' forecasting bias may not constitute a challenge if an optimal prediction would convey estimates 'near enough' to consensus forecasts. In fact, the problem of the optimal prediction in a dynamic regression framework is well-known since the introduction of Kalman (1960)'s recursive algorithm known as 'the Kalman filter' (KF, henceforth). Nevertheless, the evidence displayed in Figure 2 also contradicts this hypothesis: the Greenbook forecasts are structurally over the optimum and below the realizations (in absolute values). In other words, the optimal forecasts lie in a very strict band around zero, below (above) both the data for positive (negative) values and consensus estimate. Moreover, the existence of large outliers is ignored and, in some cases, poorly addressed: for example, the 2020 Pandemic shock in RGDP growth is interpreted as a (dramatic) increase from 0 to 5% approximately, while the corresponding observation lies at a -30 % level. Thus, the U.S. macroforecasters were effectively more accurate than an agnostic—and uninformed—econometrician.

One may be tempted to cover this lack of standard signal extraction methods by complicating the model assumed as a data generating process around which optimal forecasting exercise is applied. For example, the data may be assumed to follow a nonlinear and/or time-varying parameter evolution in several of its moments; or to be driven by some unobserved factors to be estimated using large datasets; or a combination of these features. The econometric literature is considerable and in a mature state, see Durbin and Koopman (2012) for an exhaustive introduction. However, using more complex econometric modelling, while improving our forecasting exercise, would *not* explain why U.S. professional forecasters diverge from realization of the RGDP measurements. Explaining this bias and using this information to improve our forecasting capability is the aim of this paper.

According to the literature, economic agents (and thus, professional forecasters) face some sort of limitation in their data processing capabilities and other types of costs arising from interaction with other agents (for example, reputation). These

(unobserved) costs, in turn, may easily influence the agents' utility maximization and their expectations—and consequently, their forecasts based on standard, rational expectation-based econometric models. In particular, limited attention and imperfect utility maximization justify the existence of outliers—that considerably complicate the problem of signal extraction—in forecasting. In this regard, a number of (apparent) expectation-formation biases have been documented, see Coibion and Gorodnichenko (2012, 2015) and Bordalo et al. (2020) *inter alia*.

The identification of the ultimate source of these biases is an open issue. Its exploration requires (i) a theory that explains the genesis of the macroforecasters' misbehavior and its effect on the business cycle; (ii) an estimation method that allows economists to capture the dynamics of such misbehavior. While there is a growing literature on the former issue, the latter is still at an early stage. In other words, we know that economic agents face with behavioral aspects that justify the forecasting bias but there is still not a recognized measure of aggregate distortion caused by agents' behavior. In this paper we consider macroforecater's mis-behavior an epiphenomenon of the *judgment*, defined by Svensson (2005, p. 2) as *"information, knowledge, and views outside the scope of a particular model"* and represented, by the same Author, as a Central Bank's estimate of arbitrary stochastic factor added to the econometric model equations.

Hence our research question: *How to elicit judgment directly in phase of estimation?* Our contribution is represented by the definition of a novel robust methodology to extrapolate judgment empirically. Namely, we propose the dynamically judgmental system (DJS) as a general characterization of the judgmental dynamics via state-space modelling and introduce a direct estimation method of the amount of judgment via signal extraction techniques. In our framework the (aggregate) amount of judgment in the system is parametrized by a functional of the likelihood—named "*Lq-Likelihood*", or "*Deformed Likelihood*"—of the model to be estimated. The Deformed Likelihood estimator is characterized by a parameter that defines the degree

of deformation of the logarithmic transform of the likelihood function due to the presence of additive outliers, hence providing a direct, robust measure of the degree of deformation due to judgment in the whole forecasting system. The DJS allow us to estimate the dynamics of this deformation via a peculiar version of KF, named "Deformed Kalman Filter" (DKF).

Our simulation exercise demonstrates that the DJS-DKF has good general properties in terms of accuracy in small samples and that the distribution of the deformation parameter is very well approximated by a Gaussian distribution. Finally, we apply our methodology to the Survey of Professional Forecasters of the Federal Reserve Bank (FED-SPF) with focus on the forecast of RGDP. The empirical evidence leads us to the conclusion that the judgmental dynamics varies considerably according to the ownership of the dataset and is not perfectly coincident with recession dating. These findings open new perspectives on beliefs formation in macroeconomic forecasting and related economic theory.

Section 2 allocates our contribution in the scientific debate; Section 3 describes the theoretical issues to set-up the DJS; Section 4 illustrates the robust statistical methodology to be adopted in the new framework; the results of the application on real U.S. data are illustrated in Section 5; Section 6 concludes; finally, an Appendix provides mathematical proofs, while a separate Supplement reports the results of some Monte Carlo simulations and additional results.

2 Literature

This paper contributes to both the economic and statistical literature. Concerning the economic contribution, it relates to the increasing strand of macroeconomic literature (Mankiw and Reis, 2002; Woodford, 2003; Sims, 2003) that relaxes the 'Rational Expectation Hypothesis' (REH). It also contributes to the econometric literature that relates the measurement of uncertainty and rationality testing to

professional forecasting; see Jo and Sekkel (2019) and therein mentioned literature. In particular, we are inspired by two ideas originally formalized by Townsend (1978, 1983): (i) macroeconomic forecasting is a complex activity that requires to include suitably collected experts' opinions in the classical econometric modelling of general equilibrium systems; (ii) as a consequence, it acts as a sort of clearing-market condition that, under standard REH corresponds to a Bayes-Nash-equilibrium. As a logical consequence, deviations from REH should correspond to deviations from the Bayesian updating of beliefs. This is (partially) documented by a recent strand of literature; see, *inter alia*, Manzan (2011); Ortoleva (2012); Manzan (2021); Giacomini et al. (2020); Gaglianone et al. (2020). Moreover, this paper is near to the field that reconsiders the literature in Business Cycle (and related research fields) in light of very well established results in behavioral sciences; see De Long et al. (1989); Ehrbeck and Waldmann (1996) and Ottaviani and Sørensen (2001) *inter alia*.

Secondly, we nest the literature on professional forecasters' disagreement and their learning mechanisms that complicate the definition of suitable measures of uncertainty.[2] The judgmental dynamics' effective measurement in survey data and its econometric treatment are still open issues: the literature that focuses on judgment is represented only by Manganelli (2009) and Kocięcki et al. (2012) in non-Bayesian and Bayesian frameworks, respectively, while the only reference focused on theory-based modelling is Monti (2010).

This is our methodological contribution: we provide a novel framework that allows judgment in forecasting and includes the *prequential approach* to forecast evaluation introduced by Dawid (1984)[3] and the literature on signal extraction and robust filtering: namely, we use the approach based on influence function by Masreliez and Martin (1977) to estimate the parameter that controls the degree of deformation of the likelihood function for any observation in the sample. The static version of this estimator has been introduced by Ferrari and Yang (2010). In the case of no judgment, our DKF coincides with the robust Kalman filter (RKF) by

Marczak et al. (2018) and is similar, in the economic meaning, to the one derived by Bordalo et al. (2019).[4] Finally, this paper can be viewed as a methodological counterpart of the generalized noisy information theory by Broer and Kohlhas (2022), and as a simpler alternative to the Bianchi et al. (2022)'s machine-learning technique recently introduced in the literature.

3 Theoretical framework

This section outlines the theory of DJS. Namely, Subsection 3.1 describes the formation process of judgment in a prototypical forecasting environment; Subsection 3.2 introduces the notation, while the main definitions for the set-up of the new tool are exposed in Subsection 3.3

3.1 The Judgmental Forecasting Protocol

Definition 1. (Judgmental Protocol) We assume that the forecasting bias due to judgment is the output of a repeated game (named *Judgmental Protocol*, henceforth) among three players: Policy Maker and Forecaster (henceforth PM and F, respectively) and Reality. Then, for $t = 1 \ldots, T$,

1. PM and F individually collect, process and analyze data, consisting of a set of explanatory variables, to make out-of-sample inference on a target variable; at the same time both F and PM form their own utility functions;

2. F produces (potentially biased) projections of the objective variable under alternative assumptions about the explanatory variables and/or under different scenarios;

3. PM receives the projection sets, adds judgment in each of them and, on the base of his/her objectives and information, decides which projection looks better;

4. PM announces next period's current official forecast;

5. Reality draws the realization.

Remark 1. The Judgmental Protocol above stated is a simplified version of the policy decision process illustrated by Sims (2002) and Svensson (2005). In particular, our simplifying assumptions are the following: (i) PM and F are single, homogeneous players. Thus, F may well represent both external experts and Central Bank's staff and no role is given to the Central Bank staff during the final decision process. (ii) There is no feedback effect in PM's decision—that is, Reality is not influenced by PM when the process repeats in a subsequent period, nor F replies to PM in the course of the process. In turn, this holds because we do not assume the agents' strategic behavior. (iii) Reality is neutral with respect to the decisions taken by PM and F. While this may seem counterintuitive in a purely macroeconomic policy perspective, in a microeconomic perspective it is a necessary condition to isolate the role of F and MP in the production of q time by time.

Step 1 explicitly allows the existence of the utility function and its possible divergence among PM and F. This is necessary to allow an economic interpretation of the macroforecasters' misbehavior—since it can naturally emerge directly from asymmetry in utility functions without adding any judgmental component, see Patton and Timmermann (2007) *inter alia*. The existence of (potentially different utility) functions among the two Players is definitional for setting the statistical tool that we are going to discuss in the course of this Section. Noticeably, Reality *does not* have a utility function.

Step 2 explicits two key characteristics of the judgmental process: (i) F is independent from PM in producing the projections; (ii) *projection* and *forecasting* are different objects: the former is a product of a single Player, the latter a product of a (simplistic) interaction among two Players. In Step 3 this interaction takes the form of a decision by the other Player.

Step 4 is where the judgment becomes empirically measurable. The judgmental bias is incorporated in forecasts when PM makes the decision on the best projection, which is not of public domain by definition. However, this step allows to test for the effectiveness of judgmental components only, not for the optimality of the whole forecast. The optimality testing becomes feasible only in Step 5. At this point, it is possible to compare all the phases of the forecasting process.

3.2 Notation and Set-up

We are interested in the stochastic process $[Y, X, \xi] \doteq \{[Y_t, X_t, Z_t] : \Omega \to \mathbb{R}^{k+1}, k \in \mathbb{N}, t = 1, \ldots, T\}$ where $Y_t = \{y_1, \ldots, y_T\}$ is the vector of observed data, and $X_t = \{x_1, \ldots, x_T\}'$ is a vector of explanatory variables, ξ is the vector of judgmental values added to the system and the set of all possible values taken by $[Y, X, \xi]$ is $[\mathcal{Y}_t, \mathcal{X}_t, \Xi_t]$. Moreover, it is defined on a complete probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$, where Ω is the sample space and $\mathcal{F}_t$ the event space; $\mathcal{P} \doteq \{\mathbf{p} \in \mathcal{A} : \sum_x p_x = 1\}$ defines the set of all distributions on $[\mathcal{Y}, \mathcal{X}]$ that are absolutely continuous respect to a $\sigma-$finite measure μ, $\mathcal{A}$ an algebraic subset of $[Y, X, \xi]$ representing the set of F's actions, in turn denoted as $q \in \mathbb{R}$.

The Log-likelihood function of $\hat{y}_t$ is denoted as $\mathcal{L}(\Theta, q)$; the equivalent for PM and F are $\mathcal{L}(\Theta^{PM}, q)$ and $\mathcal{L}(\Theta^{F}, q)$, respectively; the density of Y is denoted $P(Y) \doteq \int p_Y(y)dy$, where $p(\cdot)$ is a continuous density function defined on $\mathcal{L}(\Omega)$ and t is omitted to ease the notation; the (one-step-ahead) distributional and density forecasts of Y_t are denoted as $P(Y_{t+1})$ and $p(Y_{t+1})$, respectively. The utility function corresponding to the true assessment of event Y in t+1 is denoted as $U \in \mathbb{R}$ and labeled as Scoring Rule (SR). Let $\overline{\mathbb{R}} = [-\infty, +\infty]$ denote the extended real line and the functions $H(Y, X) : \mathcal{P} \to \overline{\mathbb{R}}$ and $D(X, Y) : \mathcal{P} \times \mathcal{P} \to \overline{\mathbb{R}}$ be associated with any $U(Y, X)$.

At time t, we denote $\mathcal{H}(X) = U(X, X) \equiv \sup_{X \in \mathcal{X}} U(Y, X)$ the maximum utility (or Entropy function) that a forecaster gains when X (truly) realizes and $D(X, Y) =$

$H(X) - S(X,Y)$ the Divergence function between the predictive densities of X and Y. Notice that $\mathcal{H}(X) \equiv D(X,Y)$ in case of perfect evaluation of both X and Y. Finally, a hat denotes estimates while a tilde the judgment-biased objects.

3.3 Characterizing Judgmental Dynamics

To make the previously notation operational we need the following

Assumption 1. *$\mathcal{P}$ is a strictly convex probability measure.*

Strict convexity of $\mathcal{P}$ implies that $\mathcal{U}$ is strictly proper. In turn, this requires the use of a more general distance measure compared to the one generally assumed in forecast evaluation exercises; see, among others Gneiting and Raftery (2007). The next assumption characterizes the statistical treatment of this paper:

Assumption 2. *(i) $\mathcal{U}(\cdot,\cdot,\cdot)$ and q exist; (ii) they are not observed.*

Assumption 2 (i) ensures that data are affected by bias. This assumption is supported by the literature mentioned in Section 2. Assumption 2 (ii) implies that there are at least two sources of uncertainty: the (subjective) utility function and the (objective) degree of judgment. Whereas the latter is known, it would still be possible to recover $\hat{\mathcal{U}}$ via state-space representation of the ARX(p) and to assume a quadratic loss function, so that OLS and Kalman filtering work. When the judgment term is also unknown, the classical state-space modelling is still possible, but the Kalman filter is not the minimizer of the mean square criterion because q_i is treated as a non-stochastic process that can only be inferred. The next Subsection deals with this issue.

The following definition unifies all the objects previously introduced in a common environment:

Definition 2 (Dynamically judgmental system)**.** We define *Dynamically Judgmental (Generating) System*—DJS, henceforth—the 7-ple

$\mathcal{DJS} := \{\mathcal{Y}_t, \mathcal{X}_t, \mathcal{F}_t, \mathcal{P}, S(\cdot,\cdot), H(\cdot), D(\cdot,\cdot)\}$ where $[Y, X]$ is measured by a T-dimentional dynamic system.

Example 1. (i) A p-order autoregression with explanatory variables, ARX(p), producing density forecasts $\tilde{p}(\cdot)_{t+1}$ using a T-dimensional dynamic system is a DJS, where professional forecasters' utility corresponds to $\tilde{\mathcal{U}}(\mathcal{L}, \tilde{\theta}; \cdot)$ with $\tilde{\theta} \in \Theta$ is a DJS, respectively. (ii) A p-order autoregression with explanatory variables, ARX(p), producing density forecasts $\tilde{p}(\cdot)_{t+1}$ using a static regression framework is a Judgmental System (JS) *but not* a DJS.

4 Statistical Methodology

This section illustrates the statistical treatment of DJS. Namely, Subsection 4.1 summarizes the main results by Ferrari and Yang (2010) on Deformed Likelihood estimation; Subsection 4.2 discusses the representation of the DSS in state-space form; finally, Subsection 4.3 introduces the DKF which represents the main methodological innovation and the intersection of all topics presented in previous Subsections.

4.1 Robust Estimation

The Deformed Likelihood is based on the following

Definition 3. (i) Consider any $x \in [-\infty, 0) \cup (0, +\infty]$; then its Lq-transform (or Box-Cox transform) is

$$L_q(x) = \begin{cases} \log(x) & if \quad q = 1. \\ x^{1-q}/(1-q) & otherwise. \end{cases} \tag{1}$$

(ii) Let $y_1, \ldots, y_T$ be an i.i.d. sample from $p(y_i, \theta_0)$, $\theta_0 \in \Theta$, where all the elements

are defined in Subsection 3.2. Then the Maximum Lq-Estimator (MLqE) of θ_0 is

$$\hat{\theta}_T \doteq \max_{\tilde{\theta}\in\Theta} \sum_{t=1}^{T} L_q[p(y_t;\theta)], \;\; 0 < q < 1. \tag{2}$$

(iii) Moreover, if $p(\cdot)$ is Gaussian, its Deformed Likelihood is:

$$\mathcal{L}^q(\theta; y_t) = -0.5 * [q_0 T \log(2\pi) + (\log p(y_t) + u_t^2)^q], \tag{3}$$

where: $\theta = [\mu, \sigma^2]$, $u_t = (y_t - \mu)/\sigma^2$ and π is the usual Archimedean constant; the estimated version has $\hat{q}$ and $\hat{u}_t$ instead of q and u_t.

Remark 2. The requirement that 0 < q < 1 is an identifying condition. When $q < 0$, the form of the curve changes and becomes "U"-shaped with some points coinciding with the $q > 1$ case. When $q > 1$ the function (i) tends to explode and (ii) is discontinuous around zero. This avoids the inversion of the Lq-function and its derivatives.

Remark 3. Equation (2) is the result of the maximization of the Lq-likelihood equation, which, at the t-observation, takes the form:

$$\sum_{t=1}^{T} w_t U(y_t, \theta) = 0 \tag{4}$$

which is a weighted version of the likelihood equation with $U(y_t, \theta) = p(y_t; \theta)'/p(y_t; \theta)$ and weights $w_t = p_t(y_t; \theta)^{1-q}$. When $q < 1$, data points with high likelihoods are assigned large weights. As q tends to 1, the MLqE coincides to standard Maximum Likelihood Estimator (MLE, henceforth). Typically, outliers are associated to very small weights. The estimation of (4) in a time series framework is the ultimate aim of this paper.

A small deviation from 1 is sufficient to noticeably deform the log-Likelihood; see the Supplement for a graphical view of Deformed Logarithm and likelihood functions. Thus, it is fundamental to localize the moments in which this deformation

from the LogLikelihood is most problematic. The following result goes in this direction:

Lemma 1. If $p(\cdot,\cdot)$ is Gaussian, then the estimated mean of $\hat{\theta}$ does not depend on q.

Proof. The surrogate parameter for the normal equation in a Gaussian regression is $\theta_t = (\mu', \sqrt{q}\, vech(\mathbf{\Sigma}))'$. See Example 3.2 in Ferrari and Yang. □

The parameter q is a point measure of judgment in the estimated model. Analyzing the *effects* of this deformation parameter requires to measure the discrepancy between U obtained (a) when event realizes using the same $p(\cdot,\cdot)$ or one of its moments; and (b) the maximal utility obtained by F when no bias is assumed. When this discrepancy is computable, it is possible to make comparison and hypothesis testing to verify the effectiveness of the bias, hence validate the claim on the predictive density $p_{t+1}(\cdot)$ using the empirical equivalent of (a) and (b). By Lemma 1, these are the matrices of V_t, (the optimal, non-biased variance) and $\tilde{V}_t$ (the matrices of observed, potentially biased variance). We suggest the use of the Brègman distance:

$$D(\tilde{V}_t, V_t) = \Lambda(\tilde{V}) - \Lambda(V) + \Lambda'(\tilde{V}) vech(\tilde{V}_t - V_t), \tag{5}$$

where $\Lambda : \mathbb{R}^{N\times N} \to \mathbb{R}$, with $\mathbb{R}^{N\times N}$ being the space of positive semi-definite matrices, is a scalar function three times continuously differentiable with $\Lambda(V_t) = \nabla(V_t)$ and $\Lambda'(V_t) = \nabla^2\Lambda(V_t)$, denoting the gradient and the Hessian of Λ with respect to the $K = N(N+1)/2$ unique elements of V_t and $\Lambda'(V)$ is negative semidefinite.[5] An important question is why adopt the MLqE for our problem, being this just one of the many possible families of estimators available.[6] We answer this question by using some arguments of Information Theory. Namely, the microeconomic foundation and set-up of Judgmental Protocol implies that DJS is a system characterized by Shannon Entropy, the most standard way to measure the uncertainty in complex systems.[7] The use of this kind of entropy, among the many nowadays available,

allows us to simplify our mathematical structure and to rely on standard maximization *at least* in the ideal case of no judgment. This implies that we can still rely on several standard results to make estimation, provided that we can represent a forecasting system according to a standard Entropy-based measure. This problem is nontrivial and can be solved in two steps: (i) prove that any optimal estimation in DJS leads to entropy-based methods; (ii) conversely, prove that any entropic estimation requires the DJS representation.

In the first step need to show that (i) there is a one-to-one relation between Shannon Entropy and the maximum possible utility that agents get in the forecasting system and, at the same time, between (5) and the Kullback-Liebler divergence — the most classical and simple measure of the distance among two density functions; (ii) and, as a consequence, we need to show that any minimization of (5) coincides with the minimization of the Shannon Entropy. These results are obtained by the following known preliminary results:

Lemma 2. Consider the Deformed Logarithm in (1). Then, (i) if q approaches to 1, $H_q(\cdot,\cdot)$ nests the Shannon Entropy and $D_q(\cdot,\cdot)$ the Kullback-Leibler Divergence. (ii) The minimizer over θ of $D_q(\theta,\, \theta^0)$, where θ^0 is the true parameter, is the same as the minimizer $H_q(\theta_0, \theta)$ where $q = 1/r$ and $r > 0$.

Proof. See Ferrari and Yang, pp. 755–756. □

In the second step, we need to prove that our DJS is effectively an object that nests the q-Entropy and H-function previously discussed. This is possible by the following.

Lemma 3. The q-Entropy is the H-function corresponding to the DJS.

Proof. See the Appendix. □

4.2 State-Space Representation

Until now we have introduced a set of elements for economic and statistical treatment of judgment in a forecasting system without any consideration on the dynamics. Characterizing the law of motion of such a system is the next definition's goal.

Definition 4 (State-space representation)**.** Consider the n-dimentional univariate time series y_t of forecasting outputs. Then the state-space form for y_t is the following system of equations

$$\begin{aligned}
y_t &= Z_t\alpha_t + X_t\beta + G_t\epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2); \\
\alpha_{t+1} &= T_t\alpha_t + W\beta + H_t\eta_t, \quad \eta_t \sim iid(0, \sigma_\eta^2), \\
Z &= [1,\ \mathbf{0}_{m-1}]; \qquad X = [1,\ \ \mathbf{0}_{k-1}]; \qquad H = \boldsymbol{\phi}' \\
G &= \begin{bmatrix} \boldsymbol{I}_{m-1} \\ \mathbf{0}_{m-1} \end{bmatrix}; \qquad T = \begin{bmatrix} \boldsymbol{\phi}' \\ G \end{bmatrix}; \qquad W = [\boldsymbol{\phi}',\ G]',
\end{aligned} \tag{6}$$

where Z_t is a $(m \times 1)$ vector of fixed effects, X_t a $(1 \times k)$ vector of covariates, β a $(k \times 1)$ vector of parameters, α_t an $(m \times 1)$ vector of states, T an $(m \times m)$ matrix of fixed coefficients, G an $(m \times g)$ matrix, η_t a $(g \times 1)$ vector of disturbances and $\boldsymbol{\phi}$ a $(p+1)$ vector of AR parameters. The initial conditions are:

$$\begin{aligned}
&\alpha_0 = [\mathbf{0}_m]; \quad \beta_0 = [\mathbf{0}_k]; \quad \boldsymbol{I}_{m^2} = \boldsymbol{I} \otimes [T,\ T]; \\
&H^2 = HH'; \qquad\qquad vec(P) = \boldsymbol{I}_m^{-1} H^2.
\end{aligned} \tag{7}$$

The first equation of the system (6), known as *observation equation*, has a linear regression structure where the time-varying coefficient vector α_t characterizes the strength of the process' oscillations; the second equation, known as as *state equation*, is a first order (vector) autoregression and propagates the properties of Markov-processes to the system. The initial conditions are necessary to make the estimation by Kalman Filter operational under the general assumption that some elements in the initial state $\alpha_1 \sim N(a_1, P_1)$, which distribution is independent on $\epsilon_1, \ldots, \epsilon_n$ and

$\eta_1, \ldots, \eta_n$, are unknown. Then, we call on the following assumptions:

Assumption 3. *(i)* $E(\epsilon_t, \epsilon_s) = 0$ *for all* $t \neq s$*; (ii)* $E(\eta_t, \eta_s) = 0$ *for all* $t \neq s$*; (iii)* $E(\epsilon_t, \eta_t) = 0$*; (iv)* $E(\alpha_0, \epsilon_t) = 0$ *for all* $t = 1 \ldots n$*.*

A3 (i)—(iv) are standard in the literature and set to simplify the notation and treatment. The next result provides the mathematical foundations to link (6) to the DJS:

Lemma 4. (i) The DJS is never isomorphic to JS.
(ii) The DJS is isomorphic to (6).

Proof. See the Appendix. □

This concludes the representation of the DJS via state-space modelling.

4.3 The Deformed Kalman Filter

Once the representation of the DJS via state-space modelling has been demonstrated, it is possible to study its estimation via robust signal extraction. The corresponding DKF is derived by the following

Proposition 1. Consider the system (6). Under A1 (i)–(iv) and the MLqE in (2)–(4), the DKF recursive equations are:
(i) For 1,. . . , t, . . . , n,

$$v = y_t - Z\alpha_t - X\beta; \qquad F = ZPZ' + GG'; \tag{8}$$

$$C = PZ'/F; \qquad t = v/\sqrt{q_0 F^{(q_0-1)}}; \tag{9}$$

$$v_t = \alpha + C\sqrt{F}h; \qquad P_t = \alpha_t + Cq_0FC'(h/t); \qquad Q = HG'/F; \tag{10}$$

$$P_{t+1|t} = TV_tT' + HH' - (QFQ' + QFC'T' + TCFQ')(h/t); \tag{11}$$

$$\alpha_{t+1|t} = Tv_t + W\beta + Q\sqrt{(q_0F)}(h/t); \tag{12}$$

where, for an arbitrary small number of time periods t^*,

$$h = \begin{cases} t & if \quad (t < t^*) \\ h(t, a, b) & otherwise, \end{cases} \tag{13}$$

and

$$h(t, a, b) = \begin{cases} t; & if \quad |t| \leq a \\ \frac{a}{b-a}(b - t) & if \quad a < t \leq b, \\ \frac{a}{b-a}(b + t) & if \quad -b < t \leq t - a, \\ 0 & if \quad |t| \geq b. \end{cases} \tag{14}$$

(ii) Then, by setting $I = v$ and $\sigma_I^2 = F$, we get

$$L = log(F) + log(q_0); \qquad S = v^2/F; \tag{15}$$

$$y_f = Zv_t + GG'\sqrt{q_0 F}h/(q_0 F); \qquad h_t = h/t; \tag{16}$$

$$\alpha_{t+1} = a; \qquad \Sigma_{t+1} = diag(P) \tag{17}$$

Proof. See the Appendix. □

The function h in (13) is a selector among time units that exceed a minimal value constituted by the two piecewise Hampel function $h(t, a, b)$ in (14). The latter is part of a family of functions—called *influence functions*—originally proposed in Robust Statistics to measure the dependence of an estimator on the value of any one of the points in the sample. Influence functions pre-select the path of the optimal trajectory of the (mean) estimator in a sample characterized by outliers, acting as an algorithm for automatic detection, weighting or removing outliers from a sample. This is particularly useful in long-span time series or in large databases where the identification of outliers is not trivial; see Hampel et al. (1986) among others. In our framework, the role of the Hampel function is not interesting *perse* because it serves as a "link" among the deformation parameter and the time recursions. Such a link is applied primarily via the deformed standard deviation t in (9) and the

estimated deformed state-dependent residual v_t and the deformed variance matrix P_t in equation (10); subsequently, via the conditional forecasts of the variance $P_{t+1|t}$ and the forecasted conditional states $\alpha_{t+1|t}$. The formal link between the deformation parameter and time is provided by the following

Corollary 1. *From the above recursions we get the Deformed Likelihood's weights:*

$$w_t = 1/(L)^{q_0}\mathcal{G}_t \tag{18}$$

and the averaged measures:

$$\mathcal{L} = -0.5[T\log(2\pi) + L + S]; \qquad \mathcal{L}c = -0.5[T(\log(2\pi\mathcal{S}) + 1] + L; \tag{19}$$

$$\mathcal{L}_q = 0.5((q_0 T\log(2\pi) + (L+S)^{q_0}); \qquad \hat{\mathcal{L}}_q = 0.5((\hat{q}T\log(2\pi) + (L+S)^{\hat{q}}). \tag{20}$$

Proof. Trivial if defining $\mathcal{S} = S/T$ and $\hat{q} = \sum_{1=1}^{T} h_t/T$ and $\mathcal{G}_t = -0.5((q_0 T\log(\mathcal{S}^2)\mathcal{S})$. □

Equation (18) is the core result of this paper because it defines the weight of deformation parameter q associated to each period via the gradient vector $\mathcal{G}_t$.

Finally, since Lemma 4 in the previous Subsection ensures that DJS can be represented by state-space model (6), we need to extend such a strict relationship also to their respective statistical treatments. This is possible by the following

Corollary 2. *The DKF is isomorphic with (R)KF.*

Proof. Let DJS_1 denote the DJS under Lq-transform, DJS_2 an equivalent without Lq-transformation and operator $\mathcal{T}$ the state-space system (6). By Lemma 3 and Lemma 4, it suffices to note that $\mathcal{T}$ is the same in either DJS_1 and DJS_2, since DJS_2 is a DJS_1 with q=1. □

Remark 4. The role of explanatory variables is important for the economic interpretation of the DJS-DKF. In fact, as illustrated in the next Section, they represent the

public information used by FED to make their forecasts. Thus, they are necessary to consider the parameter q as measure of PM's judgment—that is, private information as defined by Svensson (2005) and reported in the Introduction. Without X_t the model becomes a simple AR(p) model for a univariate time series and the state-space form and DKF works as well. However, q cannot be interpreted as measure of judgment but just as a tuning parameter of the Log-likelihood that summarizes the amount of deformation due to outliers—that is, something not related to private information.

5 Application

This section applies the DJS-DKF to real macroeconomic data. Subsection 5.1 describes the data; the results and their discussion are in Subsections 5.2 and 5.3, respectively.

5.1 Data

The SPF-FED, which data collection starts in 1968, is the eldest dataset on the professional forecasting activity on a macroeconomic dataset. The survey was originally administrated jointly by the National Bureau of Economic Research (NBER) and the American Statistical Association. Since the second quarter of 1990, these two institution have been replaced by the Real Time Data Research Center of the Federal Reserve of Philadelphia. In more than 50 years, the SPF-FED has considerably changed and increased, both in its amplitude (from 10 variables to forecast in the first survey in 1968, to several dozens in the current release)[8] and in its complexity (with increasing numbers and technicalities of therein items); see Croushore and Stark (2019).[9]

In particular, our research effort is stimulated by the fact that SPF-FED respondents are given a considerably short time to answer the survey.[10] Albeit assuming

that the survey has been conducted by the best human and computational capital in the market—so that publicity-seeking behavior (Laster et al., 1999) and noise-information (Orphanides, 2003), among the most popular arguments for criticism of the use of survey data in Macroeconomics, may be annihilated—the use of non-sample estimates in forecasters' output cannot be neglected completely.

This paper focuses on Real GDP growth as a case study variable due to its universal use for monitoring and addressing economic policies by deputed institutions. A set of 4 additional variables are also included to verify the behavior of the DJS-DKF with exogenous variables. These last are: production, consumption, money velocity and house prices; see the Supplement for details.[11] We apply our DSS-DKF over 10 systems of variables where all combinations are considered. All the real data—apart the SPF forecasts—has been downloaded by FRED in quarterly frequency and, whenever possible, seasonally adjusted.

5.2 Results

The estimates of the aggregated quote of judgment and the main (deformed) functionals and variance parameters are reported in the Supplement. In summary, the forecasts of the U.S. RGDP are characterized by, approximately, 20% of judgment. The first issue to address is the credibility of these results. For this purpose, we set-up a MonteCarlo experiment—also in the Supplement. According to our results, there is no evidence of a systematic over evaluation of the q-parameter. Additionally, Figure 3 plots the estimates of the corresponding dynamic system. In this exercise we make two trials, each corresponding to a different initial value of q. In both cases the estimated parameter is always 1 (that is, zero effect in judgment). Since in this case we assumed no judgment (because we only used real data) this result ensures credibility to our method.

The second issue to address is the time-varying evidence of the judgment. When is it more evident? The outputs of DJS-DKF for univariate time series, plotted in

Figure 4, panel (a), are coincident with Figure 3, panels (c) and (d), if considering only historical data. Differently, other univariate time series corresponding to SPF nowcasts and forecasts vary considerably—and in a similar fashion to Tables 4 and 5 in the Supplement. Things change when looking at nowcasts (Figure 4, panel b), where several negative peaks can be noticed at regular intervals in the first half of the sample, while in the second half, they tend to appear only in correspondence of periods of economic crisis. Instead, the one-quarter-ahead forecasts are characterized by a weighting function generally near-zero with important jumps, the frequency of which varies considerably according to the model.

The two/three/four-quarters-ahead forecasts—reported in Figure 4—are characterized by weighting function lying in 1 with a small number of downturns, generally in the first half of the sample with the only exceptions of the SARS-COV2 Pandemics where the evidence of judgment is considerable. However, such extreme changes in weighting function are characterized by a timing not consistent with what an agnostic analyst may expect: in most of the models, the evidence in support of judgment begins with a small deviation from 1 in 2001 and arrives to zero in 2005-2006 and (only in a few models for one quarter-ahead forecasts) in 2012-13. Interestingly, the official NBER recessions dates for the Great Recession are never involved.

A final issue is the effectiveness of the AR(2) estimated by MLqE with respect to the same model estimated via standard MLE. Figure 5 displays a comparisons of density estimates for the univariate time series as well as for the joint density of the system M5 (see the Supplement for details): the MLE-based estimates convey a joint density highly overlapped on one of the two explanatory variables (the industrial production in this case); on the contrary, the joint density estimated by MLqE is more centered among them. Not surprisingly, the pseudo-out-of-sample exercise exposed in Table 1 confirms that that DJS-DKF is, in, general, superior to the standard autoregression based on MLE in terms of point forecasting.

5.3 Discussion

Several considerations can be made from the above empirical investigation: first, there is a dynamics in the aggregate quote of judgment during the span of the sample, and it varies considerably if considering the two subsamples 1968-1990 and 1991-2020. Namely, the first half of the sample is characterized by an unclear prevalence of the data on non-data estimation. We interpret this finding as a consequence of three main factors: (a) the poor quality of the first decades of observations, due to the experimental state of the forecasting science and survey administration; (b) the evolution of econometric and computing techniques gradually made the use of statistical tools more affordable for economists who were asked to monitor the U.S. economy; (c) the changes in forecast making and in data collection, occurred during the two sub-samples (like, for example, the change in the GDP deflator). These factors, *per se* not sufficient to explain the judgmental dynamics' evolution, influence the utility function of the Forecaster. A change in utility function is very well-known to explain the persistence of the (non-judgmental) bias according to the standard theory, see Patton and Timmermann (2007). Thus, the problem becomes how to relate utility function to judgment and how to do testing and inference in a judgmental forecasting system (that is, how to relate the F change in utility to PMs decision). This means that the judgment becomes an indicator of a strategic behavior among the agents involved in the forecasting system. We are currently focusing on this and other related issues.

Second, the non coinciding dynamics of the switches, in the weighting function, of the majority of models with NBER recession dates implies that professional macroforecasters do not change their attitudes mechanically or according to a simple algorithm like "*use-judgment-when-recession-arrives*". Instead, several different, more complex reasoning schemes have to be assumed to justify their output. This seems consistent with Coibion et al. (2018)'s claim for a novel rethinking of the expectation formation. In this sense, the mechanics of endogenous formation of

misvaluation that can be inferred by Ilut and Valchev (2022) is confirmed by our evidence.

This finding questions the effectiveness of data revisions that FED makes periodically: if a certain amount of judgment is proved, its role in a forecast revision is not. One may argue that neutralizing the professional forecasters' judgment is exactly one of the roles of forecast revision. This consideration implies that revisions should coincide, or be in the neighborhood, with the degree of judgment found in our analysis. Instead, the revisions certified by the Bureau of Economic Analysis (BEA) in the period 1993–2019 reject this hypothesis: the role of revisions has been considerably small, only between 0.5 and 1.2 percentage points of the estimates (on average), while a simple graphical inspection of Figure 3 and Table 4 in the Supplement suggests the bias is higher.[12] A future development of our DJS methodology that also takes into account the effect of FED's staff revisions is highly recommended, specially in light of the results by Broer and Kohlhas (2022), according to which the forecasts revisions can be seen as a form of strategic behavior.

The finding that the amount of judgment is not negligible—expecially, in some periods—should not necessarily be interpreted as a claim that FED staff does not use all the available information but as the evidence that the survey respondents diversify their effort among computational mechanics and human deliberation. This interpretation is coherent with Casey (2020), according to which the three main surveys of professional forecasts are strongly driven by macroeconomic theoretical relations, so that the judgment may be seen as a link among theory and final forecasts—albeit, the recent literature suggests that professional forecasters are generally over-confident. In turn, this last finding makes us to question the nature of the link between judgment and strategic behavior. Further theoretical research will be necessary.

Finally, we are aware about the limits of our methodology. The DJS is a very general framework that applies a (time series) regression framework on data coming

from two different agents with (possibly) different utility functions. Thus, the current parametrization does not allow to discriminate the contribution of each single individual survey respondent, unless repeating the same DJS on individual data. Moreover, the Deformed Likelihood estimator, due to its logarithmic structure, may not be the best option in presence of repeated, large outliers. In this last case some extreme-value method like, for example, the one proposed by Burridge and Taylor (2006) may be preferable. Our simulations and empirical illustrations do not solve the issue of causality among judgment and the use of selected indicators. This issue requires an analytical derivation of the LqLikelihood-based estimator from a fully micro-founded forecasting scenario. While a game-theoretic foundation of forecasting has been proposed by Vovk and Shafer (2005), and is currently under development for econometric application in our current research, the causality nexus of judgment with macroeconomic variables is still an open issue.

6 Conclusions

The professional survey-based forecasts for the U.S. RGDP are characterized by a significant, persistent bias. The economic literature has explained this puzzling result by using several arguments—professional forecasters' asymmetry in utility function and the lack of agent's rationality, are among the most successfull. This paper contributes to the macroeconomic literature on macroforecasters' bias by introducing a novel judgmental approach to belief formation. The judgment, defined as everything outside the scope of the econometric model generating the RGDP forecasts, is the output of a forecasting game among three Players. The econometric counterpart of this repeated game is called Dynamic Judgmental System. For the first time in the literature, we estimates the dynamics of unobserved judgment in survey data owned by Federal Reserve Bank. This is possible by combining the properties of the Deformed Likelihood estimation and (robust) signal extraction in

a unique . The corresponding Deformed Kalman Filter allows econometricians to extrapolate the judgment from a time series of forecasting outputs using a set of explanatory variables.

Our simulation experiments reveal that the DJS-DKF does not tend to over-evaluate the judgmental quote in forecasting activity. The empirical evidence on a case study on the U.S. survey data on RGDP supports the hypothesis that professional macroforecasters have a non negligible amount of judgment which is not distributed uniformly during the span of the sample. Moreover, the knowledge of time-varying judgment is useful in applied forecasting exercises. These results confirm the recent claims by several authors that the rational-expectation framework, despite the recent refinements, should be severely modified or substituted by a more realistic hypothesis.

The proposed methodology is still in early-stages of development. Further research is required to understand the inner motivations of such a dynamics and the capability of DJS in more complex forecasting environments.

Notes

[1]Greenbook estimates is a label for what is commonly known with "consensus" forecasts.

[2]See Boero et al. (2008); Capistrán and Timmermann (2009); Patton and Timmermann (2010); Lahiri and Sheng (2010); Dovern et al. (2012); Andrade and Le Bihan (2013); Clements (2014); Rossi and Sekhposyan (2015); Andrade et al. (2016); Abel et al. (2016).

[3]Namely, our DKF is the empirical counterpart that has never developed to our best knowledge despite the Author's suggestions; see Dawid (1984), p. 289.

[4]The DKF and the one by Bordalo et al., however, do not coincide from a statistical point of view, as will be clarified in the course of the treatment. Concerning the economic foundation, since the aim of this paper is purely empirical we do not assume any particular behavioral model. The occurrence of a judgmental action is only postulated when the amount of judgment is found statistically significant.

[5]See Laurent et al. (2013) for theoretical properties of this general family of distance and its use in multivariate forecasting evaluation exercise.

[6]Among the most used in statistical literature, we mention the *Functional* Likelihood, the *Local* Likelihood and *Empirical* Likelihood approaches; see, Solari (1969); Tibshirani and Hastie (1987); Chen and Qin (1993) *inter alia*. Notice that these methods can also be combined.

[7]See Maasoumi (1993) for a compendium of Entropy-based methods for econometric applications.

[8]We refer to the Release 2020 Q3.

[9]See also the SPF website at https://www.philadelphiafed.org/surveys-and-data/real-time-data-researcl survey-of-professional-forecasters.

[10]"*Currently, the forecasters are given just over a week to send in their forecasts*" (Croushore and Stark, 2019, p. 1).

[11]We remark that the aim of this illustration is not an assessment of the monetary policy, but only an experiment of a new methodology. Evaluating judgmental issue in monetary policy would require the use of theory-based model; see Monti (2010).

[12]See the BEA website: https://www.bea.gov/gdp-revision-information.

References

Abel, J., Rich, R., Song, J., Tracy, J., 2016. The measurement and behavior of uncertainty: evidence from the ECB survey of professional forecasters. Journal of Applied Econometrics 31 (3), 533–550.

Andrade, P., Crump, R., Eusepi, S., Moench, E., 2016. Fundamental disagreement. Journal of Monetary Economics 83, 106–128.

Andrade, P., Le Bihan, H., 2013. Inattentive professional forecasters. Journal of Monetary Economics 60 (8), 967–982.

Bianchi, F., Ludvigson, S. C., Ma, S., 2022. Belief distortions and macroeconomic fluctuations. American Economic Review 105 (8), 2269–2315.

Boero, G., Smith, J., Wallis, K., 2008. Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. The Economic Journal 118 (530), 1107–1127.

Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. The Journal of Finance 74 (6), 2839–2874.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Overreaction in macroeconomic expectations. American Economic Review 110 (9), 2748–82.

Broer, T., Kohlhas, A., 2022. Forecasters mis-behavior. Review of Economics and Statistics, In press.

Burridge, P., Taylor, R., 2006. Additive Outlier Detection Via Extreme-Value Theory. Journal of Time Series Analysis 27 (5), 685–701.

Capistrán, C., Timmermann, A., 2009. Disagreement and biases in inflation expectations. Journal of Money, Credit and Banking 41 (2-3), 365–396.

Casey, E., 2020. Do macroeconomic forecasters use macroeconomics to forecast? International Journal of Forecasting 36 (4), 1439–1453.

Chen, J., Qin, J., 1993. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika 80 (1), 107–116.

Clements, M., 2014. Forecast uncertainty—ex ante and ex post: Us inflation and output growth. Journal of Business & Economic Statistics 32 (2), 206–216.

Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities? Journal of Political Economy 120 (1), 116–159.

Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review 112 (7), 2644–78.

Coibion, O., Gorodnichenko, Y., Kamdar, R., 2018. The formation of expectations, inflation, and the Phillips curve. Journal of Economic Literature 56 (4), 1447–91.

Croushore, D., Stark, T., 2019. Fifty years of the survey of professional forecasters. Economic Insights 4 (4), 1–11.

Dawid, A., 1984. Present position and potential developments: Some personal views statistical theory. The prequential approach (with Discussion). Journal of the Royal Statistical Society: Series A (General) 147 (2), 278–290.

De Long, J., Shleifer, A., Summers, L., Waldmann, R., 1989. The size and incidence of the losses from noise trading. The Journal of Finance 44 (3), 681–696.

Dovern, J., Fritsche, U., Slacalek, J., 2012. Disagreement among forecasters in g7 countries. Review of Economics and Statistics 94 (4), 1081–1096.

Duncan, D., Horn, S., 1972. Linear dynamic recursive estimation from the viewpoint of regression analysis. Journal of the American Statistical Association 67 (340), 815–821.

Durbin, J., Koopman, S. J., 2012. Time series analysis by state space methods. Vol. 38. Oxford.

Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? agency versus behavioral explanations. The Quarterly Journal of Economics 111 (1), 21–40.

Ferrari, D., Yang, Y., 2010. Maximum Lq-likelihood. The Annals of Statistics 38 (2), 753–783.

Gaglianone, W., Giacomini, R., Issler, J., Skreta, V., 2020. Incentive-driven inattention. Journal of Econometrics.

Giacomini, R., Skreta, V., Turen, J., 2020. Heterogeneity, inattention, and bayesian updates. American Economic Journal: Macroeconomics 12 (1), 282–309.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust statistics: the approach based on influence functions. Wiley-Interscience; New York.

Harvey, A., Phillips, G. D., 1979. Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika 66 (1), 49–58.

Ilut, C., Valchev, R., 2022. Economic agents as imperfect problem solvers. The Quarterly Journal of Economics forthcoming.

Jo, S., Sekkel, R., 2019. Macroeconomic uncertainty through the lens of professional forecasters. Journal of Business & Economic Statistics 37 (3), 436–446.

Kalman, R., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering.

Kocięcki, A., Kolasa, M., Rubaszek, M., 2012. A bayesian method of combining judgmental and model-based density forecasts. Economic Modelling 29 (4), 1349–1355.

Kreyszig, E., 1991. Introductory functional analysis with applications. Vol. 17. John Wiley & Sons.

Lahiri, K., Sheng, X., 2010. Measuring forecast uncertainty by disagreement: The missing link. Journal of Applied Econometrics 25 (4), 514–538.

Laster, D., Bennett, P., Geoum, I. S., 1999. Rational bias in macroeconomic forecasts. The Quarterly Journal of Economics 114 (1), 293–318.

Laurent, S., Rombouts, J., Violante, F., 2013. On loss functions and ranking forecasting performances of multivariate volatility models. Journal of Econometrics 173 (1), 1–10.

Maasoumi, E., 1993. A compendium to information theory in economics and econometrics. Econometric reviews 12 (2), 137–181.

Manganelli, S., 2009. Forecasting with judgment. Journal of Business & Economic Statistics 27 (4), 553–563.

Mankiw, N. G., Reis, R., 2002. Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. The Quarterly Journal of Economics 117 (4), 1295–1328.

Manzan, S., 2011. Differential interpretation in the Survey of Professional Forecasters. Journal of money, Credit and Banking 43 (5), 993–1017.

Manzan, S., 2021. Are professional forecasters Bayesian? Journal of Economic Dynamics and Control 123, 104045.

Marczak, M., Proietti, T., Grassi, S., 2018. A data-cleaning augmented kalman filter for robust estimation of state space models. Econometrics and Statistics 5, 107–123.

Masreliez, C., Martin, R., 1977. Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE transactions on Automatic Control 22 (3), 361–371.

Monti, F., 2010. Combining judgment and models. Journal of Money, Credit and Banking 42 (8), 1641–1662.

Orphanides, A., 2003. Monetary policy evaluation with noisy information. Journal of Monetary Economics 50 (3), 605–631.

Ortoleva, P., 2012. Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. American Economic Review 102, 2410–36.

Ottaviani, M., Sørensen, P., 2001. Information aggregation in debate: who should speak first? Journal of Public Economics 81 (3), 393–421.

Patton, A., Timmermann, A., 2007. Testing forecast optimality under unknown loss. Journal of the American Statistical Association 102 (480), 1172–1184.

Patton, A., Timmermann, A., 2010. Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. Journal of Monetary Economics 57 (7), 803–820.

Rossi, B., Sekhposyan, T., 2015. Macroeconomic uncertainty indices based on nowcast and forecast error distributions. American Economic Review 105 (5), 650–55.

Sims, C., 2003. Implications of rational inattention. Journal of monetary Economics 50 (3), 665–690.

Sims, C. A., 2002. The role of models and probabilities in the monetary policy process. Brookings Papers on Economic Activity 2002 (2), 1–40.

Solari, M. E., 1969. The "maximum likelihood solution" of the problem of estimating a linear functional relationship. Journal of the Royal Statistical Society: Series B (Methodological) 31 (2), 372–375.

Svensson, L., 2005. Monetary policy with judgment: Forecast targeting. International Journal of Central Banking.

Tibshirani, R., Hastie, T., 1987. Local likelihood estimation. Journal of the American Statistical Association 82 (398), 559–567.

Townsend, R., 1978. Market anticipations, rational expectations, and Bayesian analysis. International Economic Review, 481–494.

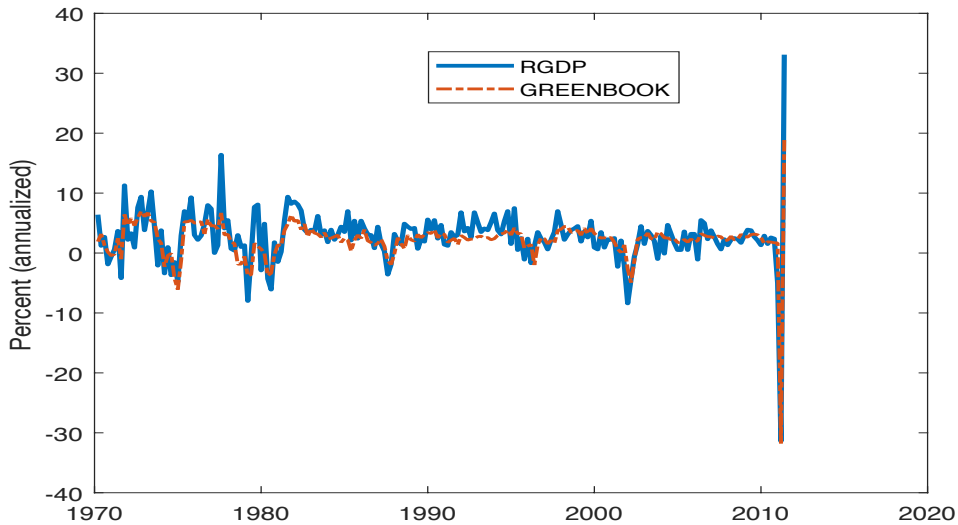
Townsend, R. M., 1983. Forecasting the forecasts of others. Journal of Political Economy 91 (4), 546–588.

Vovk, V., Shafer, G., 2005. Good randomized sequential probability forecasting is always possible. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (5), 747–763.

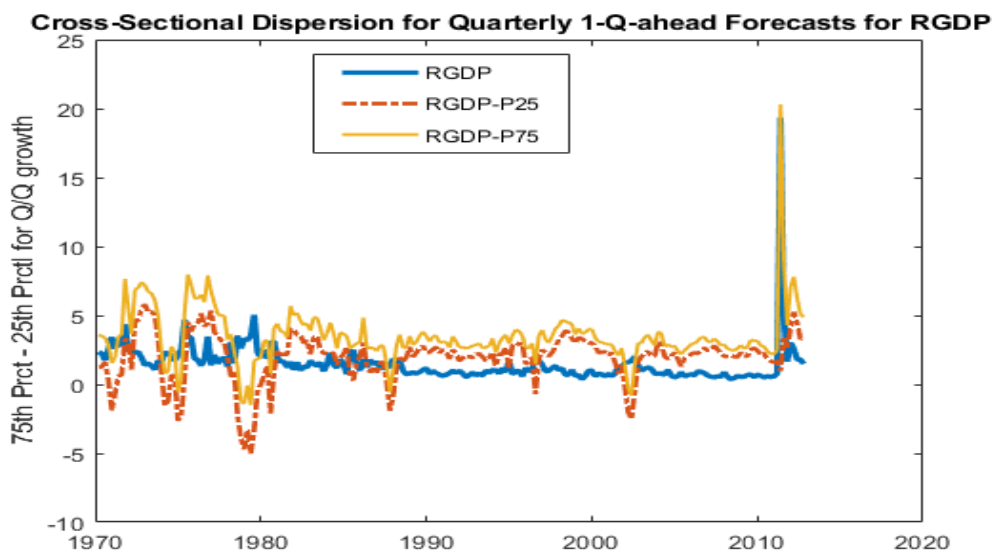
Woodford, M., 2003. Knowledge, Information and Expectations in Modern Macroeconomics: Essays in Honor of Edmund Phelps, chapter Imperfect common knowledge and the effects of monetary policy. P. Aghion, R. Frydman, J. Stiglitz & M. Woodford, Princeton: Princeton University Press.[Links], 25–58.

Figure 1: The FED-SPF forecasting bias.

(a) Aggregate bias

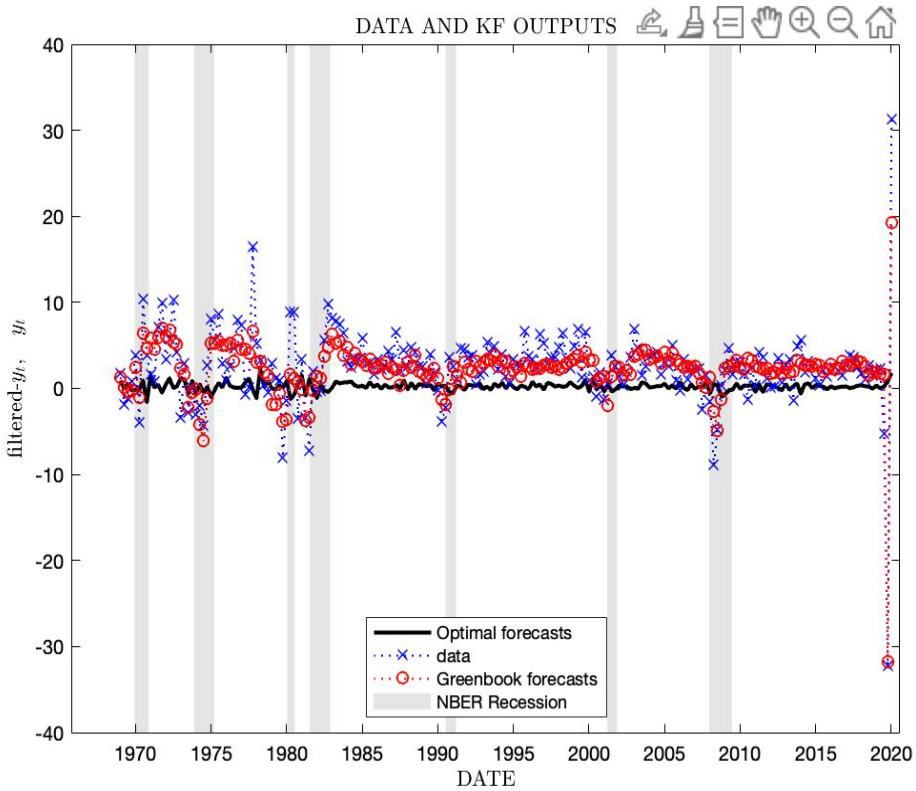


(b) Cross-sectional bias



NOTE: This figure plots the evidence of time-varying forecasting bias (top) and cross-sectional dispersion among professional forecasters (bottom). Namely, panel (a) considers the bias defined as $y_t - \hat{y}_{t+1|t}$, where y_t are realizations and $\hat{y}_{t+1|t}$ the Greenbook forecasts; panel (b) displays the $y_t - \hat{y}^{(i)}_{t+1|t}$, for $i = 1, \ldots, N$, $t = 1 \ldots, T$ with t denoting time period and i denoting the i-esim individual respondent to the survey for the 75th and 25th percentiles of the distribution of survey respondents. The sample spans from 1969Q4 to 2020Q1. All panels concern 1-quarter-ahead forecasts. The white space among 2012 and 2020 in top panels is a Matlab bag that compresses time dimension of the lines no matter of the dating labels in X-axis.

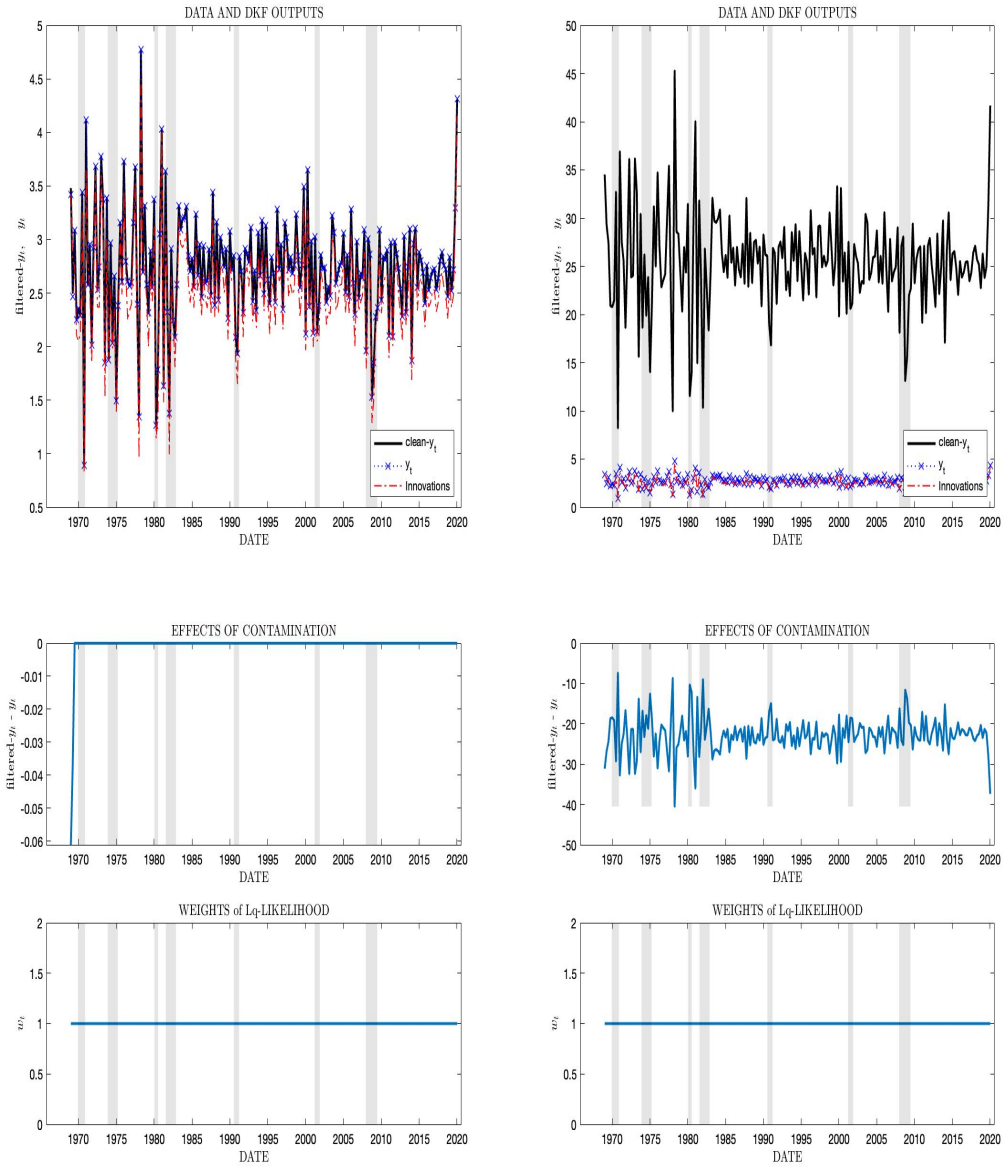
Figure 2: A comparison of optimal dynamic forecasts, Greenbook forecasts and data



NOTE: This figure compares the Greenbook 1-quarter-ahead forecasts (red) of RGDP for U.S. with optimal predictions based on the application of the Kalman filter (black) on the data (in blue). Grey areas are the NBER Recession dates.

Figure 3: Estimation of unobserved judgment on U.S. Real GDP using historical data only

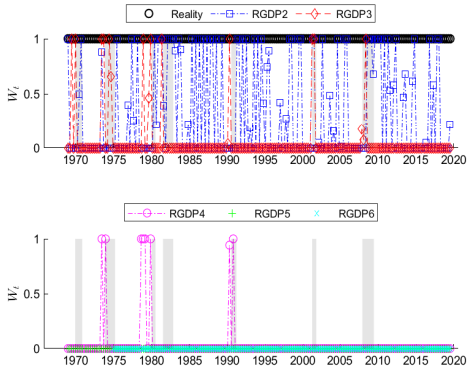
(a) $q_0 = 1$ **(b)** $q_0 = 0.1$



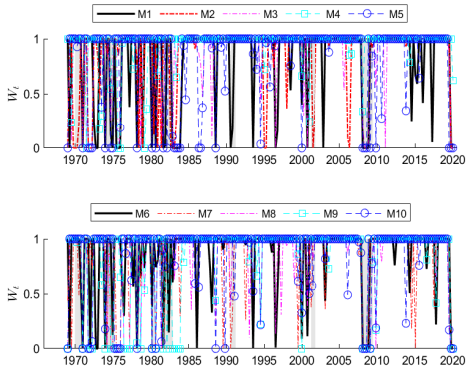
NOTE: This figures plots the results of the application of the DKF on the data of U.S. Real GDP by the Federal Reserve Bank of St. Louis. The left panels deal with $q_0 = 1$ assumed in the DJS, while the right panels display the results of a DJS with $q = 0.1$. The top panels display the time series of the observations, the series of estimates as well as the vector of innovations; the bottom panels explicit the effect of the DKF: namely, central sub-panels display the filtered series and the bottom sub-panels plot the weighting function (18).

Figure 4: Estimation of unobserved judgment on U.S. Real GDP forecasts using several DSS specifications

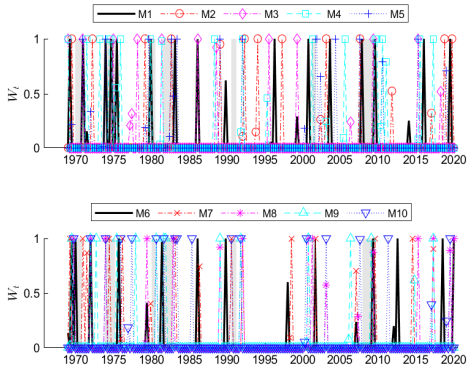
(a) Historical data



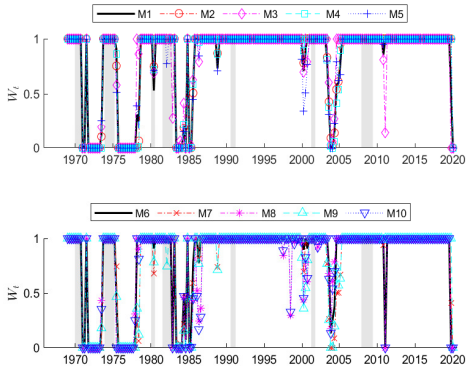
(b) Nowcasts



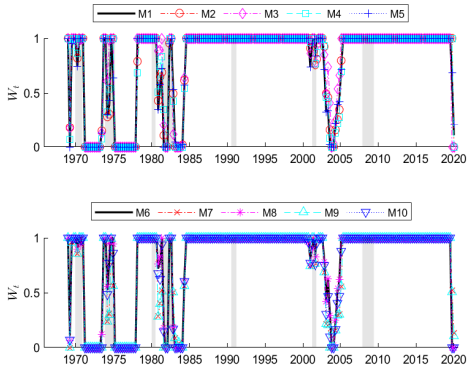
(c) 1-quarter-ahead



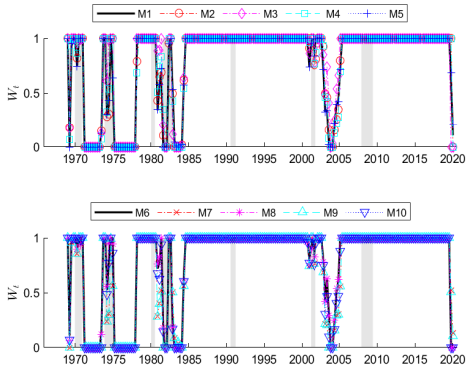
(d) 2-quarters-ahead



(e) 3-quarter-ahead

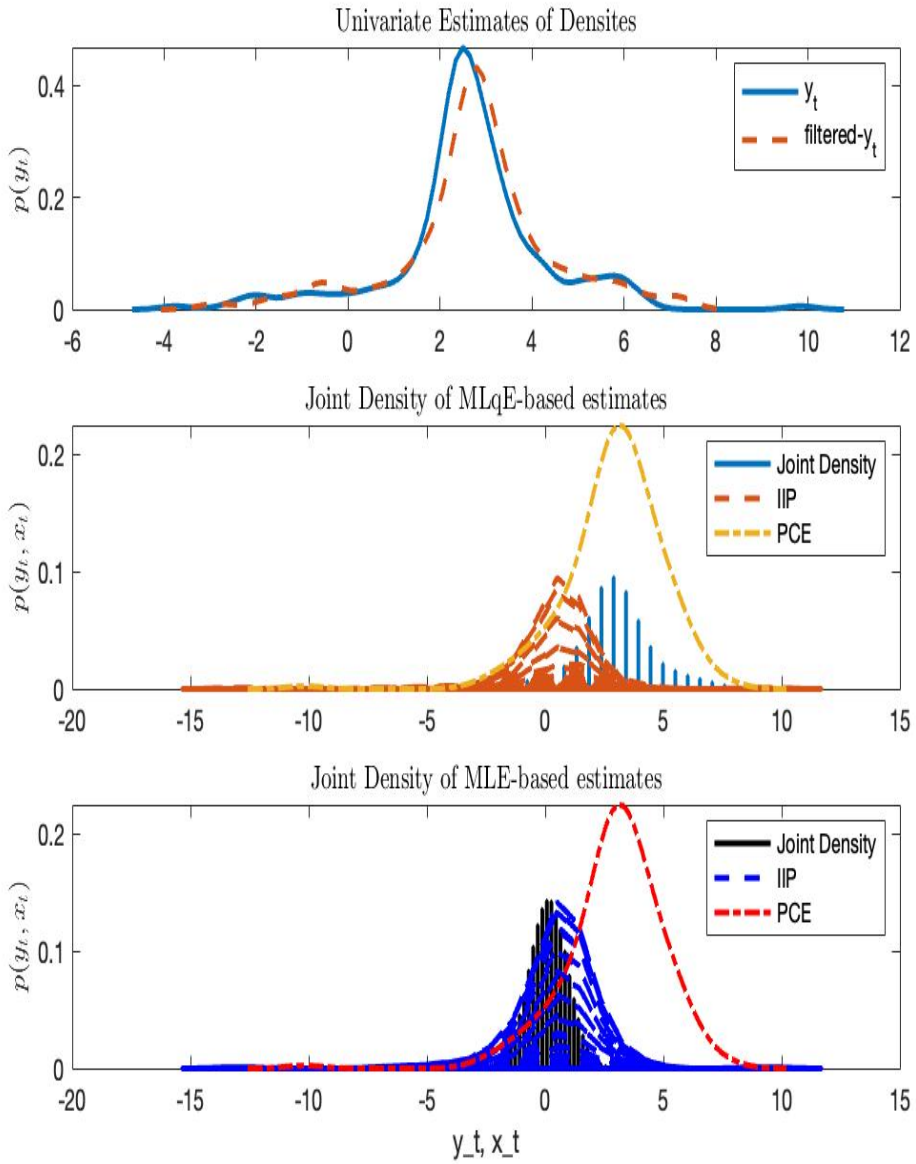


(f) 1-year-ahead



NOTE: This figure plots the results of the application of the DJS on U.S. Real GDP forecast by SPF using several data and model specifications.

Figure 5: Comparing MLE and MLqE in Density Estimation



NOTE: This figure plots (in the upper panel) the estimates of univariate probability density functions of data and the DKF-estimates of M5 (a system with one-year-ahed RGDP forecasts as dependent variables and IIP and PCE as exogenous variables; in the central panel, the univariate density functions of IIP and PCE as well as their joint density function by MLqE; and, in the lower panel, the univariate density functions of IIP and PCE as well as their joint density function by MLE. See the Supplement for further data explanation and labels.

Table 1: In-samples and out-of-sample goodness-of-fit of DJS-DKF

Model	Maximum Likelihood Estimation					
	Reality	Nowcasts	1-q-ahead	2-q-ahead	3-q-ahead	1-yr-ahead
RMSE	2.1703	3.3934	0.1702	0.2208	0.2746	0.6805
"In-sample RMSE"	1.2400	1.1449	0.0902	0.1044	0.1303	0.3502
"Out-of-sample RMSE"	1.2551	1.7220	0.0643	0.1205	0.1299	0.3730
	Maximum LqLikelihood Estimation					
RMSE	1.9257	2.7796	0.1681	0.2208	0.2746	0.6800
"In-sample" RMSE	1.1745	1.0043	0.0640	0.1035	0.1300	0.3500
"Out-of-sample RMSE"	1.2250	1.6881	0.0533	0.1192	0.1295	0.3765

NOTE: This table reports the RMSE of the AR(2)-DJS estimated using standard Kalman Filter (upper part) and the DKF (lower part) for the case of the whole sample as well as for the case of pseudo-out-of-sample exercise with fixed window; in this last case, the estimation part ("in-sample") is 1970:Q2–1990:Q1, while the evaluation part ("out-of-sample") is 1990:Q2–2020:Q1.

A Appendix

A.1 Proof of Lemma 1

To prove an isomorphism among two sets it suffices to show that these two differ at most by the nature of their elements; see Kreyszig (1991), among others. DJS and JS do not differ by the nature of the observations enclosed in the 7-ple $\{\mathcal{Y}_t, \mathcal{X}_t, \mathcal{F}_t, \mathcal{P}, S(\cdot,\cdot), H(\cdot), D(\cdot,\cdot)\}$, but in the way in which the collection of $[\mathcal{Y}, \mathcal{X}]_t$ is estimated—by dynamic system, the former and by static regression, the latter. On the contrary, the DSS and (6) may differ at most by the nature of the observations in $\mathcal{Z}_t$.

A.2 Proof of Lemma 4

It descends directly from (a) the part (ii) of Lemma 1 and (b) Lemma 2. Lets consider each of them separately.

(a) the part (ii) of Lemma 1 is necessary but not sufficient. If DJS were not isomorphic to the state-space form in equation 6, there could exist an H-function different from H_q that maximizes U(X,Y) given q. On the opposite side, assume there exists a H-function different from H_q that maximizes U(X,Y) given q; in this case, there is no implication of uniqueness of q nor $H \neq H_q$.

(b) Lemma 2 is a sufficient, non-necessary condition for Corollary 1 to hold. Part (i) ensures a one-to-one relation among q-Entropy, a known D-function and the DJS; part (ii) ensures the one-to-one relationship among the true parameter θ_0 and the amount of the aggregate quote of judgment represented by q.

A.3 Proof of Proposition 1

(i) CASE 1: q=1. By Lemma 3, the Lq-Likelihood function coincides with the standard Likelihood; by Corollary 2, the DKF coincides with the RKF. Thus, the proof is delegated to Kalman (1960); Duncan and Horn (1972); Harvey and Phillips

(1979); Marczak et al. (2018).

CASE 2: 0 < q < 1. Since $q \neq 1$ implies a bias to $\hat{\alpha}$ and $\hat{\beta}$, these are no more the minimum mean square error of α and β. Thus, we only need to justify the introduction of (a) $t = v/\sqrt{q_0 F^{q_0-1}}$ in eq. (10)–(12); (b) q_0 in the second addend of (11).

Lets start from (b). By Lemma 2, we only need to notice that, in the Harvey and Phillips (1979) notation here adopted, the variance is parametrized by F via P_t, and that F is a scalar. Thus, there is no need of half-vectorizing P. The same argument holds for equation (13).

To prove (a), let us remark that, in exponential family, $\theta^* = \theta_0/q$, where $\theta = [\mu, \sigma^2]$. By Lemma 2, μ is not influenced by q, thus it suffices to consider σ^2/q, where σ^2 is known. Thus, the only unknown variable is $1/q$. According to the normal equation (4), the optimal solution is the first derivative of $\sqrt{F^q}$.

(ii) Direct consequence of (i).

SUPPLEMENT
to
"Can we estimate macroforecasters' mis-behavior?"

Emilio Zanetti Chini

University of Bergamo
Department of Economics
Via dei Caniana, 2 - 24127, Bergamo (ITALY)
e-mail: emilio.zanettichini@unibg.it

December 2022

Abstract

This Supplement includes further results that corroborate the treatment of the Main Document. Namely, Section 1 includes some basic results which knowledge is assumed in the Main Text; Section 2 reports the Monte-Carlo simulation of the DJS; finally, Section3 reports details of estimation.

1 Preliminary Theory

Proposition 1. If $U(X) = A\log(P(X;\theta))B(\theta)$ - where A and B are an arbitrary constant and function of θ, respectively, then

(i) the maximal utility reduces to a Shannon's Entropy, that is:

$$\mathcal{H}(X) = -\mathcal{E}[\log p(X)], \tag{1}$$

which is also called *expected score* and

(ii)

$$\begin{aligned}\mathcal{D}(p(X), p(Y)) &= -\mathcal{E}\left[\log\left(\frac{p(X)}{p(Y)}\right)\right] \\ &= -\int_{\Omega} \log\left(\frac{p(X)}{p(Y)}\right),\end{aligned} \tag{2}$$

that is the Kullback-Liebler divergence

Proof. (i) is a result of the geometric interpretation of a decision problem by Schervish (1989).

(ii) is a well-known result. □

The gaussianity is a requirement to use KL as information criterion. Adding judgment leads to misspecification, hence to an inconsistency of the utility function. In statistics, this is equivalent to say that the SR is not proper, see Gneiting and Raftery (2007).

2 Simulation

This section investigates the empirical properties of the DJS-DKF in a MonteCarlo simulation exercise. Namely, Sub-section 2.1 describes the Data Generating Process (DGP) adopted; Sub-section 2.2 reports the results; finally their relevance is discussed in Sub-section 2.3.

2.1 The Data Generating Process

We consider two different DGPs:

$$y_{1,t}^{(i)} = 1.24 y_{1,t-1}^{(i)} - 0.68 y_{1,t-2}^{(i)} + \epsilon_{1,t}^{(i} + O_t, \quad \epsilon_{1,t}^{(i)} \sim N(0,1) \tag{3}$$

and

$$y_{2,t}^{(i)} = 1.24y_{2,t-1}^{(i)} - 0.68y_{2,t-2}^{(i)} - 1.4x_{t,1}^{(i)} - 0.88x_{t,2}^{(i)} + \epsilon_{2,t}^{(i)} + O_t, \quad \epsilon_{2,t}^{(i)} \sim N(0,1), \quad (4)$$

where, in both (3) and (4),

$$O_t = (I_{t=45})85\sigma\epsilon_t + (I_{t=130})30\sigma\epsilon_t \quad (5)$$

defines two innovation outliers taking value 1 at the 45th and 130th observation and zero otherwise, $i = \{1, \ldots, I\}$ denoting the i-th draw of the process $\{y_t\}_{t=1}^T$ with a total number of draws $I = 5,000$ and the length of the of the two outliers (of different strength) has been set only for exposition issue. Eventually, O_t will be deleted for illustrative reasons, so that the $y_{1,t}^{(i)}$ (henceforth "DGP 1") and $y_{2,t}^{(i)}$ (henceforth "DGP 2") become a pure autoregression and an autoregression with exogenous variables, respectively.

We recommend to pay special attention to the role played by who looks at the analysis: in this case, (3) represents what professional forecasters observe and analyse (the autoregressive model is her subjective choice). More in detail, $y_{1,t}^{(i)}$ is a linear autoregressive model with highly stationary behavior, which allows us to focus on the effects of outliers and the initial value of the deformation parameter q_0. This may be represented by a macroeconomic indicator that is affected by an unexpected shock that pervades the time series dynamics. On the other hand, $y_{2,t}^{(i)}$ describes a mixed scenario: in addition to the initial judgment q_0, the autoregression is spurred by an exogenous variable x_t. If y_t is assumed as a time series of announcement by forecast evaluators, the entire DJS can be interpreted as a fully dynamic system where the forecaster's output (in this case, x_t) is an input that co-exists with Reality. The discrepancy represents the basis for an ex-post assessment by the forecast user via utility function. This last is incorporated in the estimation step via the Lq-Likelihood. Hence, q_0 also represents a sort of a-priori of forecasts' evaluators with

respect to which forecasters may adjust their forecasts.

In our exercise, we consider three cases $q_0 = \{0.1, 0.9.1.0\}$, corresponding to high, low, and no initial judgment; and three sample sizes $T = \{50, 100, 200, 500\}$ corresponding to a very small, small, medium and large-sized samples, respectively. Clearly, when the size of the sample is short, the second innovation outlier is not considered.

2.2 Results

A draw of the simulated AR(2) process in equation (3) without outlier is displayed in Figure 3, while the same process with additive outliers can be verified in Figure 4. In the former, when $q_0 = 1$, we can notice a sort of initialization effect both in effects $(y_f - y)$ and in the weighting function w_t, being the only deviation (from zero for the effects, from 1 for the weights) in the first three observations. In the latter, the weighting function is always one apart the observations coinciding with outliers. The weighting function is also an indicator of the strength of the innovation outliers. Differently, there is not a large difference among processes with and without outliers when high judgment is assumed initially, but oscillations among the two extremes are more frequent in the case with outliers.

The simulated distribution of the deformation parameter q is shown in Figure 5. In general, the q-parameter is almost normally distributed, according to the assumptions on errors in the DGP. When $q << 1$, the mean is upward biased of a decimal, approximately and, when assuming exogenous regressors, there is a small increase in the right tail. When no judgment is assumed, the parameter is almost completely concentrated around 1.

The average measures described in equations (20) and (21) of the Main Text and the estimated distortion parameter computed from the complete MonteCarlo exercise are reported in Tables 2 – 3. Several facts can be noticed: first, the distortion parameter is near to the unit when the initial value q_0 is high; in this case, there

is no significant difference among the two DGPs, despite the presence of outliers. When q_0 is low, the estimated q is instead almost uniformly near 0.20; few exceptions are due to low sample sizes. Innovation outliers tend to increase this estimate of approximately 50%.

Second, the presence of innovation outliers blows up the difference among the estimated Lq-Likelihood and the Lq-Likelihood under q_0; for example, consider the case of pure AR process with $T = 100$ in table 3, where $\mathcal{L}_{q0} = 3,875$ and $\mathcal{L}_{\hat{q}} = 9,357$ (more than double). Such an inflation effect is generally more pronounced in the case of a high initial q_0 and lower for a low initial value of the same parameter.

Third, the predicted error variance (PEV) as well as the standard error of regression tends to increase as q_0 diminishes. However, the proportionality of this error variance inflation is not linear if innovation outliers spur the process. For example, consider the case of T=50: without outliers, the PEV goes, approximately, from 1.4 to 2.6; with one outlier, it rises from 320 to 1,460 (that is, a completely different order of magnitude).

2.3 Discussion

Despite the simplicity of the DGP assumed in our simulation exercise we can draw some important conclusions regarding the introduction of a judgmental component in a time series process: first, any judgment does not modify the mean significantly. This observational result complements the theoretical analogue in Ferrari and Yang (2010)[1]. However, it should not be confused with the simulated effects of the judgment, which are evident since the resulting time series is another autoregression with different variance.

Second, and consequently, a large judgmental bias has long-run effects in the variance, and specially if the process is characterized by outliers. This is perfectly in line with the theoretical fundaments by Ilut and Valchev (2022) on the dynamics

[1]See, in particular, *ibidem*, page 759 and 769.

of the cost of the deliberation. In fact, according to these authors, the corresponding true policy function used by forecasts' evaluators in their evaluation is unknown. This reflects in uncertainty on the true SR, which can only be inferred by forecasters by bayesian methods to update their beliefs. Thus, they gradually accumulate information about the optimal quotation as function of the underlining state. Such an accumulation causes the judgment's propagation in the entire forecasting process.

Third, fixing an initial amount of q_0 does not guarantee that the distribution of that parameter has a mean coinciding with that initial value, apart an almost coincidence in the case of no judgment assumption ($q = 1$). This seems desirable because it avoids any automatic determinism when applying the filter, allowing the possibility to have an estimate of the amount of judgment that may diverge from q_0 in any new recursion.

Finally, this simulation exercise enlightens on the connections among our DJS approach and the classical assessment methods like the probability integral transform introduced in Economics by Diebold et al. (1998), which can be re-interpreted as follows: if the (true) DGP is associated to a true initial assessment of the quote of judgment (q_0), the forecaster does not use learning so that her beliefs (corresponding to the weighting function) do not need to change.

A similar scenario holds in case of outliers (that is, Figure 4), where the only discrepancies from 1 in the weighting function coincide with the timing of outliers. Thus, under normality and perfect specification, there is a perfect learning and immediate beliefs updating. On the other side, when some amount of judgment is assumed, the resulting signal generates a noise that adds to the forecasting process forcing the forecasts' evaluator to frequently deliberate about her forecasts. In turn, this reflects in frequent changes in weighting function; in this last case, the outliers do not modify the general logic apart an increase in the magnitude of the change in forecast due to subjective judgment. According to Tables 2 and 3, the inclusion of exogenous regressors – which represents information by other agents in the system or

economic indicators – does not allow us to modify this interpretation of the graphical 'one-sided' results.

3 Estimation Results

When the DJS-DKF is applied on 'one-side' real data of RGDP (that is, the data without foreecasters' outputs as exogenous variable, see Table 4 first column of top panel), the estimated aggregate judgment is zero (equivalent to $q = 1$), so that the DJS-DKF coincides with a standard State-Space form of an AR model estimated by Marczak et al. (2018)'s RKF (that is, by MLE).

On the opposite side, when real data are substituted by nowcasts and forecasts (from second to last column of the upper panel) things change radically, because $\hat{q} = 0.32$ (that is more than two thirds of the estimates are due to the judgment) for nowcasts and arriving to zero for long-run horizon forecasts. This means that these long-run horizon forecasts are purely judgmental.

Such an extreme variation among the estimates based on historical data and forecasts may be easily explained by an omitted variable bias. Thus, we replicate our estimation exercise with all the systems defined in the Supplement. The results are reported in the second panel of the same Table. Noticeably, the estimated quote of judgment is always minimal: the $\hat{q}$ is always upper than 0.99 and, in one case—M4, a system with house prices as only exogenous covariate—zero, since the estimated Log-likelihood coincides with the deformed Likelihood of the estimated model. This is a more realistic scenario, albeit the ARX-DJS of this peculiar case is based only on historical data and not on forecasts. This means that whenever $\hat{q} \neq 1$, the DJS is characterized by (minimal) standard misspecification. Interestingly, the same scenario holds for results in the third panel, corresponding to nowcasts. Thus, nowcasting is not affected by expectation-formation bias.

The estimates resulting from the ARX-DJS based on one-quarter-ahead forecasts

(fourth panel) are instead characterized by a non negligible amount of judgmental bias: the estimated q is never less than 0.80—this time without differences among the different models. Similar results are reported in Table 5, with possibly stronger evidence of judgmental bias in higher horizons. In general, Models M5 and M9 seem the most expectation formation-biased for two-quarter-ahead forecasts, while M5, M7 and M10 are in three-quarter ahead, being the estimated $\hat{q}$ lower than 0.80. The results for the one-year-ahead forecasts are similar—apart from the fact that all the deformation parameters are above the same thresholds.

References

Diebold, F., Gunther, T., Tay, A., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39 (4), 863–883.

Ferrari, D., Yang, Y., 2010. Maximum Lq-likelihood. The Annals of Statistics 38 (2), 753–783.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Ilut, C., Valchev, R., 2022. Economic agents as imperfect problem solvers. The Quarterly Journal of Economics forthcoming.

Marczak, M., Proietti, T., Grassi, S., 2018. A data-cleaning augmented kalman filter for robust estimation of state space models. Econometrics and Statistics 5, 107–123.

Schervish, M., 1989. A general method for comparing probability assessors. The annals of statistics 17 (4), 1856–1879.

Tables and Graphs

Table 1: Definitions

Variables		
Label	Definition	FED code
y_t	Real GDP	A191RL1Q225SBEA
ip_t	Index of Industrial Production	IPB50001SQ-PCH
c_t	Real Personal Consumption Expenditure	DPCERO1Q156NBEA
m_t	Velocity of M2 aggregate	M2V-PCH
h_t	Average Sales Price of Houses Sold, (in percent change from 1 year ago)	ASPUS-PC1

Systems of variables		
Label	Endogenous Variable	Exogenous variables
M1	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	ip_t
M2	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	c_t
M3	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	m_t
M4	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	h_t
M5	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, c_t]$
M6	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, m_t]$
M7	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, h_t]$
M8	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, c_t,\, m_t]$
M9	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, c_t,\, h_t]$
M10	$\{y_t,\, y_t^N,\, y_{t+1},\, y_{t+2},\, y_{t+3},\, y_{t+4}\}$	$[ip_t,\, c_t,\, m_t\, ,\, h_t]$

NOTE: This table reports the definitions of the models adopted in Section 5 of the Main Text.

Table 2: Simulation of the DJS

# obs.	**Parameter**	DKF with q_0=.9		DKF with q_0=0.1		KF	
		AR	ARX	AR	ARX	AR	ARX
	$LogLik.$	-81.6883	-70.2086	-77.8280	-57.0563	-80.9066	-72.1364
	$LqLik_{q0}.$	-64.6703	-57.7557	-5.3523	-5.2564	-	-
T=50	$LqLik_{est}.$	-9.0161	-69.2499	-4.7618	-10.9022	-	-
	$ConcLogLik.$	-79.6058	-69.5059	-79.6081	-53.6853	-79.6001	-71.5500
	PredErrVar	1.4548	1.0274	2.6591	2.4096	1.3578	1.0075
	σ	1.4649	1.0240	2.2335	1.5145	1.3578	1.0045
	$\hat{q}$	0.9849	0.9952	0.0880	0.2179	-	-
	$LogLik.$	-154.1712	-135.2904	-191.954	-115.2804	-164.7407	-143.1970
	$LqLik_{q0}.$	-97.4076	-91.1888	-10.0338	-9.9047	-	-
T=100	$LqLik_{est}.$	-152.7386	-131.4826	-22.2903	-20.6313	-	-
	$ConcLogLik.$	-156.2843	-134.2817	-144.2997	-83.5018	-161.0238	-142.5610
	PredErrVar	1.4427	1.0509	4.4611	2.4621	1.4367	1.0080
	σ	1.4463	0.9091	3.0783	1.5383	1.4367	1.0034
	$\hat{q}$	0.9950	0.9894	0.2247	0.2129	-	-
	$LogLik.$	-268.2211	-276.8432	-233.2850	-232-4844	-274.5920	-285.8236
	$LqLik_{q0}.$	-215.9634	-220.5399	-19.1704	-19.1590	-	-
T=200	$LqLik_{est}.$	-266.4892	-274.1103	-37.7171	-40.1943	-	-
	$ConcLogLik.$	-267.9858	-276.1221	-222.2339	-220.0763	-274.0324	-285.2241
	PredErrVar	0.9316	1.0320	2.4233	2.4909	0.8979	1.0140
	σ	0.9330	1.1045	1.5465	1.5581	0.8979	1.0080
	$\hat{q}$	0.9452	0.9963	0.1984	0.2114	-	-
	$LogLik.$	-715.8411	-690.781	-668.5619	-583.5638	-717.6153	-713.7228
	$LqLik_{q0}.$	-550.8887	-538.6978	-46.8613	-46.8113	-	-
T=500	$LqLik_{est}.$	-696.8721	-683.7737	-80.9314	-98.2274	-	-
	$ConcLogLik.$	-714.3682	-689.9998	-607.4176	-553.1045	-717.9314	-713.0975
	PredErrVar	1.1076	1.0300	3.1497	2.5263	1.0285	1.0105
	σ	1.1128	1.0209	1.8711	1.5681	1.0285	0.9940
	$\hat{q}$	0.9906	0.9964	0.1730	0.2103	-	-

NOTE: This table reports the results of a Monte Carlo simulation of the AR(2) in equation (3) for 5,000 replications.

Table 3: Simulation of the DJS with innovation outliers

# obs.	**Parameter**	DKF with q_0=.9		DKF with q_0=0.1		RKF	
		AR	ARX	AR	ARX	AR	ARX
	$LogLik.$	-8,185.1595	-3,541.5624	27,988.1695	-1123.0180	-24,361.9007	-3,546.5756
	$LqLik_{q0}.$	-3,127.5315	-1,392.1680	-6.0865	-5.57458	-	-
T=50	$LqLik_{est}.$	-6,749.2277	-2760.3855	-27.3896	-12.2314	-	-
	$ConcLogLik.$	-214.6680	-167.5190	-218.2628	-116.6820	-243.2484	-169.8284
	PredErrVar	322.3633	145.4756	1,467.9250	72.0643	321.9795	140.9056
	σ	325.6053	139.8505	1118.8163	44.1481	321.9795	40.9056
	$\hat{q}$	0.9800	0.9722	0.3017	0.2123	-	-
	$LogLik.$	-10,325.2642	-3,709.3747	34,350.7486	-3,709.3747	-30,716.4690	-3,274.1722
	$LqLik_{q0}.$	-3,875.1017	-1,474.7892	-10.7119	-1474.7892	-	-
T=100	$LqLik_{est}.$	-9,357.3144	-3,140.6201	-59.6578	-3240.6201	-	-
	$ConcLogLik.$	-404.4254	-301.7293	-409.1442	-301.7293	463.7831	-306.6698
	PredErrVar	203.7122	72.2785	685.8146	72.2785	612.4712	72.2874
	σ	204.7385	72.4129	686.3635	72.4129	612.4712	72.2874
	$\hat{q}$	0.9900	0.9847	0.3576	0.3550	-	-
	$LogLik.$	-660.0193	-3941.8992	-1,801.0973	-1344.1019	-1,311.6599	-1344.1019
	$LqLik_{q0}.$	-405.2554	-1621.8979	-19.5007	-19.3966	-	-
T=200	$LqLik_{est}.$	-635.0304	-3443.6199	-61.2487	-40.7272	-	-
	$ConcLogLik.$	-433.1544	-565.8913	-45.5808	-378.4371	-527.0222	-378.4372
	PredErrVar	4.8290	35.7801	20.2384	11.2707	22.5612	20.2384
	σ	4.8470	11.2707	12.6720	11.2707	23.2240	20.2384
	$\hat{q}$	0.9926	0.9850	0.3020	0.2090	-	-
	$LogLik.$	-2,586.0027	-4,551.02422	-7,947.1910	-1751.9711	6,423.9046	-4578.1265
	$LqLik_{q0}.$	-1,335.6732	-1992.2934	-47.2546	-46.9886	-	-
T=500	$LqLik_{est}.$	-2,514.2672	-4254.1267	-150.7868	-98.6685	-	-
	$ConcLogLik.$	-1,223.4403	-1229.5731	-1,281.3423	-813.2265	1,503.4601	-1254.6905
	PredErrVar	8.5851	16.5283	63.0292	14.5502	23.8537	16.4699
	σ	8.6023	16.4565	31.1006	14.8759	23.8537	16.4699
	$\hat{q}$	0.9960	9.992	0.3073	0.2095	-	-

NOTE: This table reports the results of a Monte Carlo simulation of the AR(2) in equation (4) for 5,000 replications.

Table 4: Application of DJS on a real data

Parameter	AR-DJS					
	Reality	Nowcasts	1-q-ahead	2-q-ahead	3-q-ahead	1-yr-ahead
LogLik.	-468.4657	-1489.8564	-461.9765	-390.8810	-392.9121	-402.3322
LqLik_{q0}.	-	-19.9360	-19.7775	-19.7496	-19.7506	-17.6498
LqLik.	-468.4657	-67.7087	-13.0732	-7.9296	-0.5000	- 0.5000
ConcLogLik	-220.5766	-376.5798	-279.5885	-276.4205	-285.5864	-247.0211
PredErrVar	5.0348	15.7099	2.4397	19.7496	0.0000	0.0000
Var	4.0348	14.5380	4.2190	7.9296	3.2130	4.1234
$\hat{q}$	1.0000	0.3252	0.0654	0.0387	0.0000	0.0000

ARX-DJS using historical data

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-1623.2627	1229.7122	-1300.3216	-505.3844	-1844.8413	-1685.1902	-1621.0797	-1775.5874	1943.2856	-1937.1170
LqLik_{q0}.	-817.6411	-655.3521	-684.9145	-	-9069405	-842.7295	-816.7548	879.1666	-946.2225	-943.7677
LqLik.	-1515.3346	-1177.3059	-1258.8399	-	1718.3826	-1572.1119	-1513.3328	-1662.6731	-1808.5075	-1802.8617
ConcLogLik	-555.7900	-522.7840	-529.9145	-401.6643	-570.5677	-560.1369	-555.6334	-566.1715	-546.5120	-576.1494
PredErrVar	16.6411	10.8688	13.0184	3.1493	19.2962	17.4463	16.7035	18.4998	20.4369	20.3654
Var	13.9402	10.1459	10.8290	3.1493	16.0900	14.5475	13.9814	15.4209	17.0412	16.9816
$\hat{q}$	0.9903	0.9934	0.9952	1.0000	0.9903	0.9903	0.9903	0.9910	0.9903	0.9903

ARX-DJS using nowcasts

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-817.9064	-692.9322	-796.8510	-365.5465	-824.4603	-854.8055	-817.2072	-854.8797	-846.3158	-878.4762
LqLik_{q0}.	-478.7125	-423.0148	-469.4129	-	-481.6017	-494.9354	-478.4042	-494.9680	-491.2111	-505.2948
LqLik.	-774.2543	-674.1225	-775.4747	-	-780.3141	-809.0816	-773.6080	-808.4216	-800.5094	-830.2176
ConcLogLik	-472.1178	-449.3252	-468.6414	-343.9949	-473.1769	-477.9421	-472.0043	-477.9535	-476.6308	-481.5129
PredErrVar	10.9522	5.2518	10.5864	1.7910	11.037	11.5545	10.9094	11.5549	11.4081	11.9592
Var	6.1588	4.9539	5.9554	1.7910	6.2222	6.5154	6.1521	6.5161	6.4333	6.7441
$\hat{q}$	0.9903	0.9948	0.9952	1.0000	0.9903	0.9905	0.9903	0.9903	0.9903	0.9903

ARX-DJS using 1-q-ahead forecasts

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-695.5482	-694.3590	-710.9412	-695.2588	-697.4238	-715.5222	-699.1189	-725.3174	-698.7149	-726.4419
LqLik_{q0}.	-20.0207	-20.0205	-20.0237	-20.0207	-20.0211	-20.0246	-20.0214	-20.0265	-20.0213	-20.0267
LqLik.	-302.5698	-306.3062	-312.2619	-304.1970	-296.5920	-320.8958	-297.6357	-302.2374	-294.7472	-299.8278
ConcLogLik	-259.509	-259.3844	-261.6703	-259.5102	-259.8130	-262.2916	-260.0480	-263.6152	-259.9926	-263.7622
PredErrVar	6.5185	6.5122	6.6834	6.5126	20.0211	6.7472	20.0214	6.7653	6.5082	6.7660
Var	7.1845	7.1730	7.3332	7.1817	296.5920	7.3775	297.6535	7.4721	7.2150	7.4829
$\hat{q}$	0.8209	0.8243	0.8258	0.8223	0.8155	0.8315	0.8161	0.8150	0.8137	0.8128

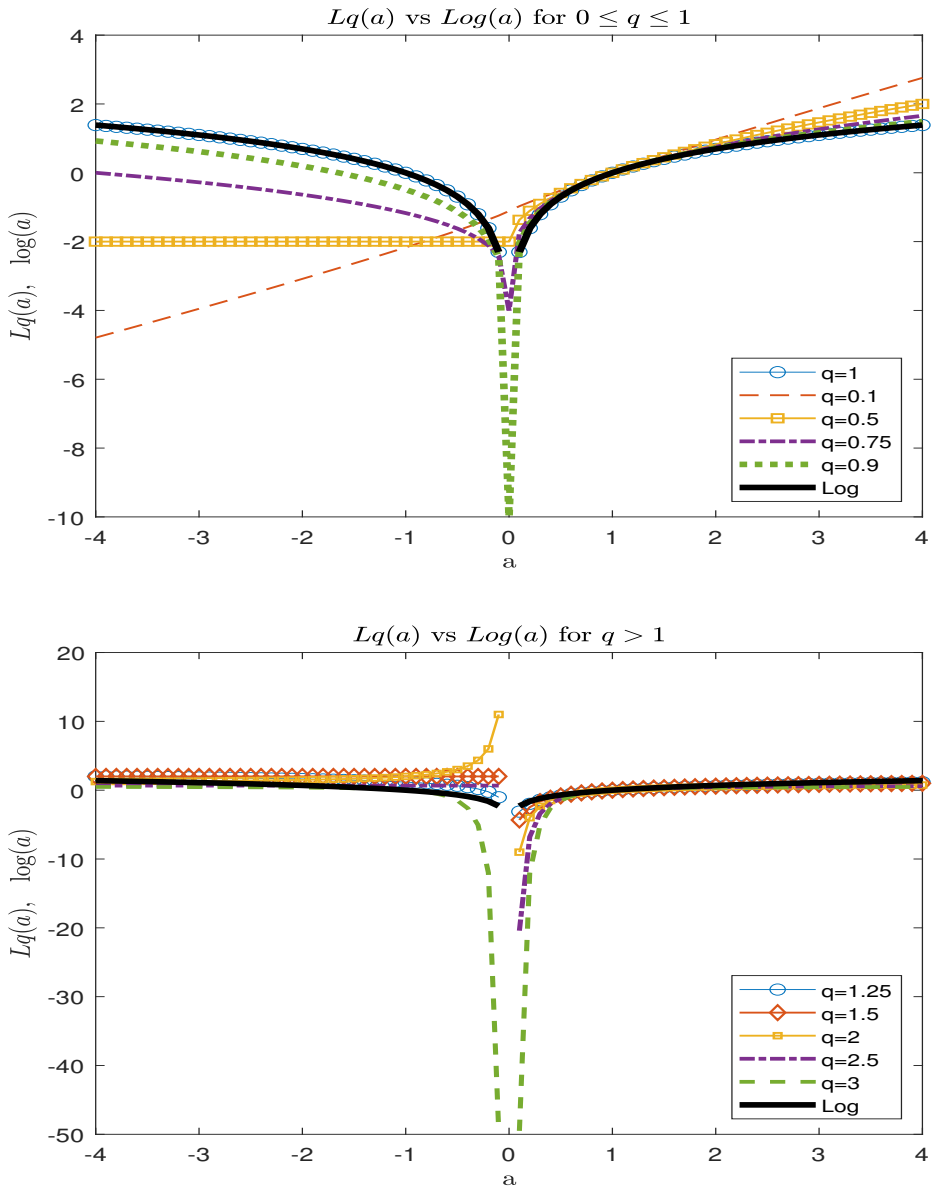
NOTE: This table reports the estimates of the DJS-DKF exposed in Sections 3 and 4 of the Main Text. The top panel concerns about pure Autoregressive case (that is, the application of the DJS-DKF to the single time series of real data, nowcasts and forecasts by SPF; the further three panels concerns the systems with exogenous variables described in Supplement. In upper-middle, lower-middle and bottom panels, the dependent variables are the historical value, the nowcast and 1-quarter-ahead forecast of RGDP, respectively.

Table 5: (Continue...)

Parameter	ARX-DJS using 2-q-ahead forecasts									
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-690.2583	-684.7648	-694.3104	-685.8272	-690.4604	-694.1474	-695.8906	695.8567	-692.4070	-697.6114
LqLik_{q0}.	-20.0197	-20.0186	-20.0205	-20.0188	-20.0197	-20.0204	-20.0208	-20.0208	-20.0201	-20.0211
LqLik.	-278.4897	-295.3149	-289.0269	-290.6663	-277.2926	-288.6965	-284.0565	-282.5177	-274.4097	-280.4103
ConcLogLik	-258.8548	-258.0744	-259.4173	-258.2263	-258.8836	259.3947	-259.6396	-259.6355	-259.1575	-259.8804
PredErrVar	6.4104	6.3991	6.4856	6.3915	6.4055	6.4831	6.4802	6.4742	6.4127	6.4803
Var	7.1329	7.0799	7.1271	7.0901	7.0138	7.1705	7.1873	7.1870	7.1536	7.2039
$\hat{q}$	0.8007	0.8169	0.8095	0.8127	0.7995	0.8093	0.8048	0.8034	0.7964	0.8011
	ARX-DJS using 3-q-ahead forecasts									
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-681.6467	-680.2310	-677.7409	-697.7192	-681.4417	-678.6237	-687.7169	-677.4012	-678.3060	-687.9334
LqLik_{q0}	-20.0179	-20.0176	-20.0171	-20.0175	-20.0179	-20.0173	-20.0192	-20.0171	-20.0173	-20.0193
LqLik.	-265.1522	-270.5129	-283.0759	-271.2372	-265.9414	-277.2797	-259.3687	-283.8229	-277.4724	-259.0032
ConcLogLik	-257.6943	-257.4895	-257.1265	-257.4162	-257.6654	-257.2561	-258.5572	-257.0775	-257.2107	-258.5879
PredErrVar	6.2796	6.2725	6.3483	6.2713	6.2789	6.3388	6.3089	6.3483	6.3320	6.3098
Var	7.0490	7.0354	7.0115	7.0305	7.0471	7.0199	7.1076	7.0082	7.0169	7.1097
$\hat{q}$	0.7892	0.7949	0.8074	0.7957	0.7901	0.8018	0.7821	0.8081	0.8020	0.7817
	ARX-DJS using 1-yr-ahead forecasts									
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
LogLik.	-363.0309	-363.3719	-354.7980	-354-1534	-362.5196	-363.5108	-351.8168	-362.5816	-370.7473	-370.7661
LqLik_{q0}	-17.8160	-17.8161	-17.8120	-17.8117	-17.8157	-17.8162	-17.8106	-17.8158	-17.8195	-17.8159
LqLik.	-201.3670	-201-6545	-202.2749	-201.8684	-201.3274	-200.0900	-202.3765	-200.5811	-200.0486	-199.1993
ConcLogLik	-186.4725	-186.4578	-184.5411	-184.3938	-186.3556	-186.5831	-183.8407	-186.3710	-188.2270	-188.2309
PredErrVar	4.1001	3.9738	4.0973	3.8957	4.0892	4.1482	-4.0709	4.1378	4.2139	4.2265
Var	4.4009	4.4049	4.3120	4.3050	4.3954	4.4061	4.2797	4.3960	4.4846	4.4848
$\hat{q}$	0.8135	0.8139	0.8193	0.8189	0.8137	0.8109	0.8210	0.8123	0.8075	0.8059

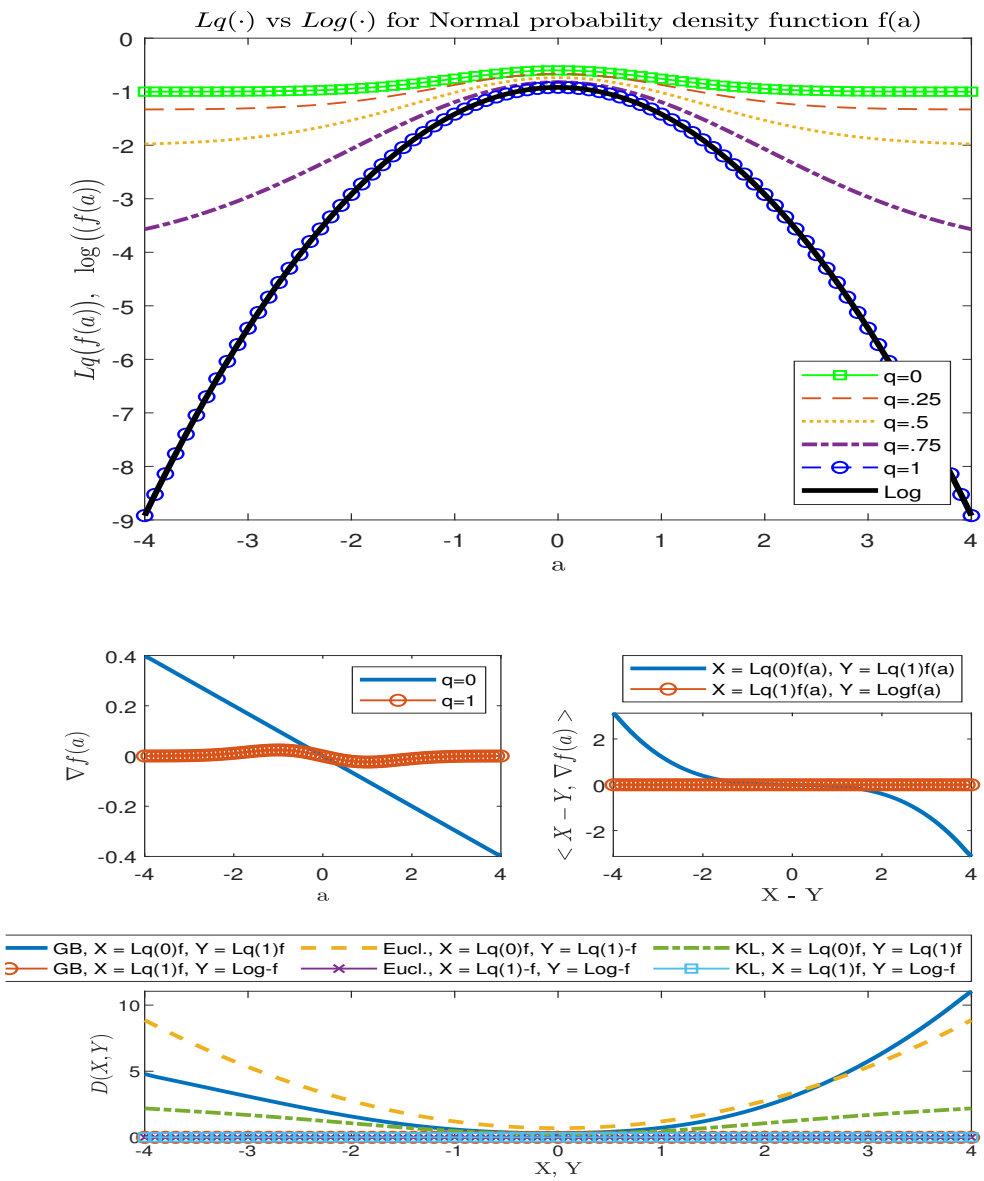
NOTE: This table continues to report the estimates of the DJS-DKF exposed in Sections 3 and 4 of the Main Text. In this case, the same systems with exogenous variables described in Supplement are analyzed, but the dependent variable is a forecast with medium/long horizon.

Figure 1: The Deformed Logarithm function



NOTE: This figure displays the Lq-function applied to the numerical sequence $a = [-4; 4]$ for different values of q and compares it with the natural logarithm function.

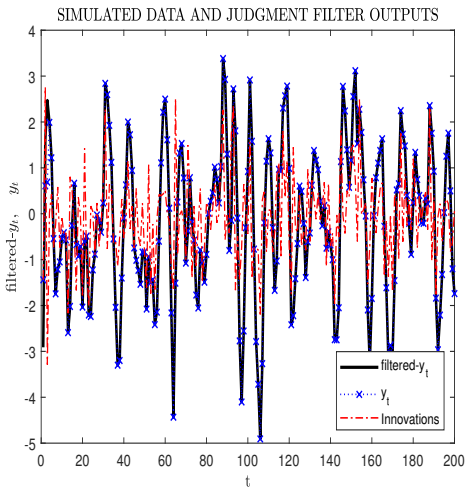
Figure 2: Functional analysis of the Deformed Logarithm



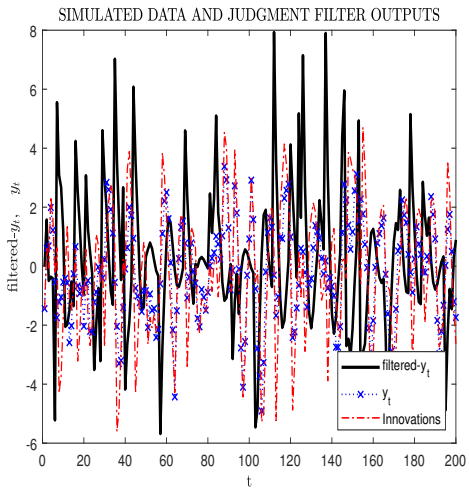
NOTE: This figure displays the behavior of several functionals of the Deformed Logarithm. The upper Panel shows the Lq-function of a standard normal probability density function f over the sequence $[-4; 4]$ given a set of values of q. The lower Panel plots (i) the Lq-function over the gradient vector of a for $q = 1$ and $q = 0$ (high-left sub-panel); (ii) the inner product among $X - Y$ and the gradient of the probability density function of y, both in the case of $X = Lq(f)$, with q=0 and $Y = Lq(f)$ with q=1 (blue), and in the case that $X = Lq(f)$, q=1 and $Y = \log(f)$ (red-circles), in high-right sub-panel; (iii) several Divergence functions (namely, the Generalized Brègman, Euclidean and Kullback-Liebler obtained for different combinations of $Lq(f)$ for the same couple $q = \{0, 1\}$ and their comparisons with $\log(f)$ (bottom sub-panel).

Figure 3: Application of the Deformed Kalman Filter on simulated data.

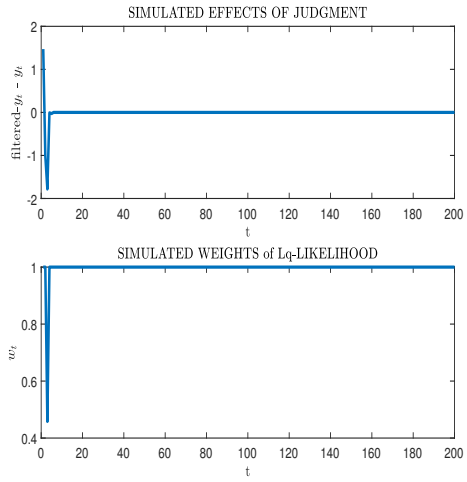
(a) The simulated process with $q_0 = 1$



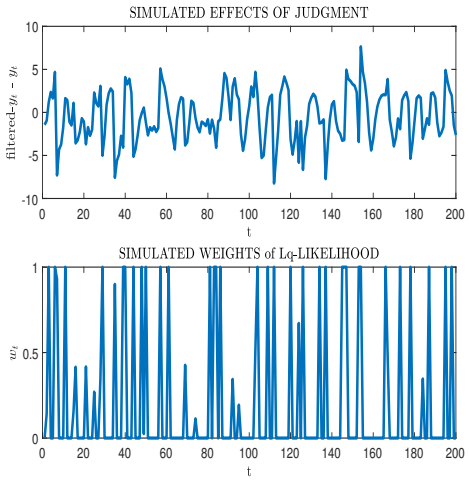
(b) The simulated process with $q_0 = 0.1$



(c) Filter components when $q_0 = 1$



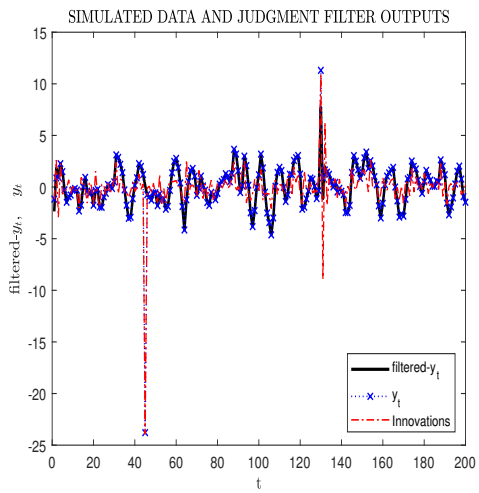
(d) Filter components when $q_0 = 0.1$



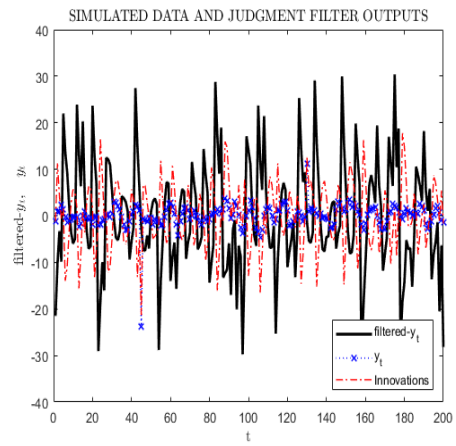
NOTE: This figure plots the result of a judgment filtering exercise using simulated AR(2) process described in (3) without innovation outliers. The upper panels display the original data, the "clean" process and the innovations. The lower panels display the output of the spread clean process vs the same data and the estimated weights of the LqLikelihood. Left panels deal with $q_0 = 0.1$, while the right panels report the results for a process assumed having $q = 1$.

Figure 4: Application of the Deformed Kalman Filter on simulated data with innovation outlier.

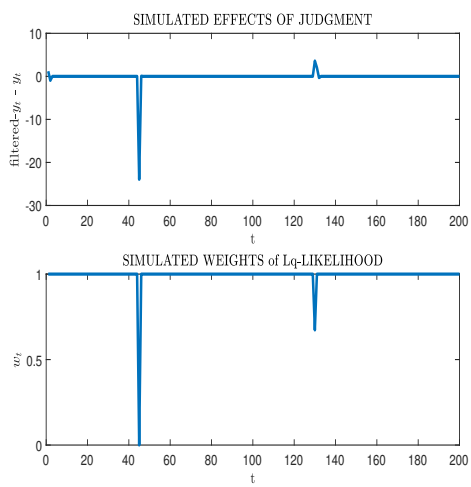
(a) The simulated process with $q_0 = 1$



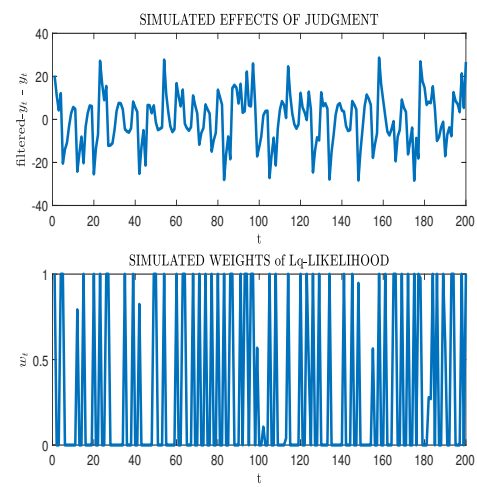
(b) The simulated process with $q_0 = 0.1$



(c) Filter components when $q_0 = 1$

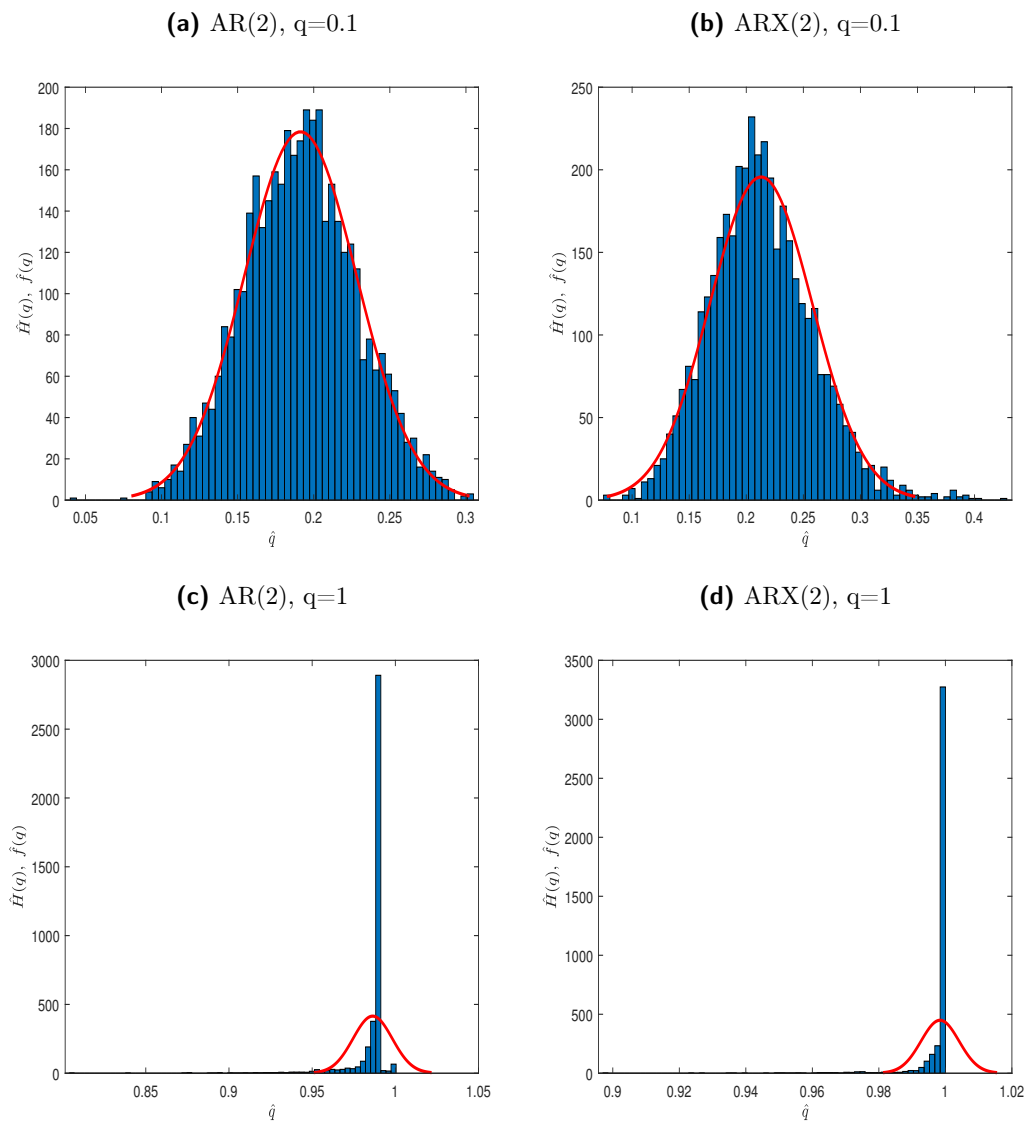


(d) Filter components when $q_0 = 0.1$



NOTE: This figure plots the result of a judgment filtering exercise using simulated AR(2) process with innovation outlier described in (3) – (4). The upper panels displays the original data, the "clean" process and the innovations. The lower panels display the output of the spread clean process vs the same data and the estimated weights of the LqLikelihood. Left panels deal with $q_0 = 0.1$, while the right panels report the results for a process assumed having $q = 1$.

Figure 5: The simulated distribution of the judgmental bias parameter q



NOTE: This figure displays the histograms of the estimated p resulting from the Monte-Carlo exercise in Section 2. Namely, the higher panels concern the results for $q = 0.1$ and the lower ones the results for $q = 1$; left panels deal with pure autoregressive case, while the right ones include exogenous regressors.