UNIVERSITÀ DEGLI STUDI DI BERGAMO

Department of Economics

WORKING PAPERS

Measuring Unobserved Strategic Judgment

Emilio Zanetti Chini

November 2021 - WP N. 6 Year 2021

Measuring Unobserved Strategic Judgment

Emilio Zanetti Chini*

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

November 2021

Abstract

We estimate the dynamics of unobserved strategic judgment in macroeconomic forecasts via state-space methods. This is possible by using a new, micro-founded framework named Dynamic Scoring Structure (DSS), in which judgment arises as a deformation of the Log-Likelihood function of the estimated forecasting model and extrapolated via robust signal extraction. The properties of the new methodology are investigated via MonteCarlo simulation. An application to the survey forecasts of Real GDP of U.S. economy suggests that judgment has a dynamics related but not coincident with Business Cycle phases. The impact of this finding is also discussed.

Keywords: Deformed Likelihood, Disagreement, Evaluation, Filtering, Inattention, Repeated Games.
JEL: C12, C22, C44, C53.

*The Author thanks B. Annicchiarico, Alessandra Luati, and Tom Stark for their feedbacks. He is also grateful to all participants to 41 (Virtual) International Symposium on Forecasting 2021, the internal Seminar Series in Mathematics and Statistics of University of Bergamo. Finally, the Author is in debt with doctors, nurses and workers of Policlinico "S. Matteo" of Pavia, without whose (free) 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 human science in any historical time. When the Italian composer formulated his statement, the Scientific Knowledge was a privilege for a restricted élite and the circulation of informations dramatically slower than today. Thus, despite its humor, the statement should be considered as a (partial) truth. On a different hand, when the Nobel Laureate was in activity, the technological progress was already pervasive and the development of statistical tools for data processing had a so rapid acceleration that the incentive for exploiting the capability of data mining was – for the first time – superior to the incentive to think new approaches to the Economic Science, hence explaining his paroxysmal statement.

These epiphenomena – costly understanding and the information excess/over-easy-manipulation – are antithetic only apparently. In fact, both of them may induce the investigator to a bias, which nature varies if these are considered singularly. This paper aims to enlighten the differences among these two aspects of the modern economic research. The most recent tendencies in economic literature (Varian, 2014) suggest that the increased availability of ultra-high-dimensional datasets makes the need of an empirical strategy for the second aspect (that is, a correct and efficient selection of the available amount of information) prevails on the need of dealing further with the first aspect, like investigating theoretical issues – albeit there is not consensus on the best methodology to use for data-rich environments.[1] Not strangely, the use of private information channels by the economic agents and their role in belief formation is still under debate by theorists, and, *a fortiori*, its correct

[1] Only recently some important arguments in favor of 'dense'-families of models has been proved by Giannone et al. (2021).

assessment is a primary objective in econometric forecasting. We contribute to this issue by assuming a forecasting scenario in which an economic agent that makes forecasts (or forecast producer – FP henceforth) co-exists and interacts with an other agent that uses and evaluates it (forecast user – FU, henceforth); not secondly, we assume that this interaction produces effects over time. One of them is the rise of estimation obtained by extra-sample information (or *judgment*, henceforth), arising as consequence of a mis-evaluation by both the agents, Thus, judgment is a subjective variable that justifies the (non infrequent) situations in which FPs do not behave according to the Bayes rule without invoking behavioral arguments.

This paper proves that this scenario can be formalized by state-space modelling, and that the amount of forecast error due to judgment introduced by agents with a strategic perspective (or *strategic judgmental bias* – SBJ, henceforth) can be estimated via signal extraction techniques. Thus, we introduces the Dynamic Scoring Structure (DSS, henceforth), a peculiar state-space representation of an autoregressive model with exogenous covariate that enable us to incorporate judgment in FP's and FU's outputs, where the (aggregate) amount of strategic judgment in the system is parametrized by a functional of the likelihood – named "*Lq-Likelihood*", or "*Deformed Likelihood*" – of the model to be estimated. This last is governed by a parameter that defines the degree of deformation of Likelihood due to the presence of additive outliers, hence providing a direct, robust measure of the quote of judgment in the FP/FU's model. Specularly, the FP may consider the same deformation parameter as a sort of *a priori* for the possible mis-evaluation of her forecast by FU. Thus, the DSS allows econometricians to measure the SJB as an endogenous equilibrium of a repeated game of two agents and the reality and to estimate its dynamics via a peculiar version of the Kalman (1960)'s classical recursive algorithm, named "Deformed Kalman Filter" or "Judgment Filter" (JF, henceforth).

Our simulation exercise demonstrates that the DSS-JF has good general properties in terms of accuracy in small samples and that the distribution of the deforma-

tion parameter is very well approximated by a Gaussian distribution. Finally, we apply our methodology to the Survey of Professional Forecaster of Federal Reserve Bank (FED-SPF) with focus on forecast of Real GDP. Our results make us conclude that the SJB has a dynamics that varies considerably according to the ownership of the dataset and is not perfectly coinciding with recession dating. These findings open a new perspective in macroeconomic forecasting and related economic theory on the belief formation.

The rest of the paper is organized as follows: the next Section 2 allocates our paper in the scientific debate; Section 3 describes the DSS-JF; the results of the application on real U.S. data are illustrated in Section 4; finally Section 5 concludes. A separate Supplement provides preliminary theory and MonteCarlo simulation results.

2 Literature

Since the Seventies of the past Century a consensus among economic scientists, inspired by contributions by R. Lucas, T. Sargent and C. Sims, has grown around econometric models for general equilibrium based on rational expectation hypothesis (REH), according to which all economic agents uses perfectly their available information no meaning about exogenous shocks. These models has been proved sufficiently general and flexible to measure and explain many economic issues; see, among others, Canova (2011) for survey.

Due to the increasing criticism,[2] the mainstream macroeconomic literature has relaxed the axioms of REH since a couple of decades; see, among others, Mankiw and Reis (2002); Woodford (2003); Sims (2003). Contemporaneously, the research effort in econometric methodology has allowed the development of models that accounts for several sources of complexity, so leading to important improvements in the measurement of uncertainty and rationality; see Jo and Sekkel (2019) and therein

[2]See Conlisk (1996) for a survey.

literature.

This paper contributes to this last strand of literature by investigating the use of judgment by economic agents asked to produce and evaluate forecasts. In particular, we are inspired from two ideas originally formalized by Townsend (1978, 1983): (i) macroeconomic forecasting is a complex activity that requires to complement the classical econometric modelling of general equilibrium systems with the opinions of experts, collected in form of survey; (ii) as a consequence, it acts as sort of clearing-market condition that, under REH, corresponds to a Bayes-Nash-equilibrium. Secondly, we are stimulated by the theoretical finding by Ottaviani and Sørensen (2006) that forecast competitions, if not properly set, lead to a strategic behavior of the same FP, who has an improper incentive to announce a forecast different from the 'true' one – that is, the output of an optimization on sample data.

The literature on professional forecasters' disagreement, its links to uncertainty, as well the learning mechanisms of the same FPs that complicate the definition of suitable measures of uncertainty, is large.[3] However, the effective measurement of the judgment dynamics on survey data and its econometric treatment are still open issues: the literature that focus on FPs is represented only by Manganelli (2009) in a non-bayesian, and Kocięcki et al. (2012) in bayesian framework, respectively, while the only reference focused on FU is Monti (2010). To our best knowledge, no literature is available if considering strategic interaction among FUs and FPs.

We mind this gap by providing a new statistical framework that englobes the *prequential approach* to forecast evaluation introduced by Dawid (1984)[4] and the literature on optimal and signal extraction and robust filtering: in the case of no SJB, the JF coincides with the filter derived by Marczak et al. (2018), who modify the diffuse Kalman Filter (De Jong, 1991) allowing for a rescaling via influence function

[3]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).

[4]Namely, our JF is the empirical counterpart that has never developed to our best knowledge despite the Author's wish; see *Ibidem*, p. 289

as suggested by Masreliez and Martin (1977). Namely, we use the influence function to estimate the deformation parameter that characterize the Deformed Likelihood estimator introduced by Ferrari and Yang (2010) for any observation in the sample. The new filter is able to describe the output of a repeated game. Moreover, this paper is a development of Zanetti Chini (2019), who use a static regression approach.

Our research applies (parts of) the theory by Ilut and Valchev (2020) that consider (inattention-driven) bias an effect of a costly deliberation; secondly, it can be considered an empirical counterpart of (i) the literature that studies how the utility functions that allow FUs to assess the credibility of FP can be tested via non-bayesian methods (Dekel and Feinberg, 2006; Al-Najjar and Sandroni, 2014; Pomatto et al., 2014) and (ii) the theoretical literature (Vovk and Shafer, 2005) that considers the forecasting process as an output of a game among several economic agents;[5] finally, it offers an alternative explanation to the evidence of failure of bayesian approach to belief-updating explained by Manzan (2011); Giacomini et al. (2020); Manzan (2021), among others.

3 Theoretical framework

3.1 Notation

We are interested in the stochastic process $Z \doteq \{Z_t : \Omega \to \mathbb{R}^{k+1}, k \in \mathbb{N}, t = 1, \ldots, T\}$. This process is partitioned as $Z_t \equiv [Y_t, X_t]$, 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 exogenous predictors and the set of all possible values taken by Z_t is $\mathcal{Z}_t \equiv [\mathcal{Y}_t, \mathcal{X}_t]$. Moreover, it is defined on a complete probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$, where Ω is the sample space; the event space $\mathcal{F}_t$ is partitioned as $\mathcal{F}_t \equiv [\Pi, \Psi] \in \mathbb{R}^k$ to denote the sub-spaces of FU and FP, respectively; $\mathcal{P} \doteq \{\mathbf{p} \in \mathcal{A} : \sum_x p_x = 1\}$ defines the set of all distributions on $\mathcal{Z}$ that are absolutely continuous respect to a $\sigma-$finite measure μ, $\mathcal{A}$ an algebraic

[5] See also Olszewski (2015) for a survey.

subset of Z representing the set of FP/FU judgmental actions, in turn denoted as $\pi \in \mathbb{R}$.

The Log-likelihood of FU and FP are denoted, respectively, $\mathcal{L}(\Pi)$ and $\mathcal{L}(\Psi)$, while the density of Z (or each of its partitions) is denoted $P(Z) \doteq \int p_Y(z)dz$, 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 Z_t are denoted as $P(Z_{t+1})$ and $p(Z_{t+1})$, respectively. Then, there exists an utility function $U \in \mathbb{R}$ corresponding to the true assessment of event Z in t+1; this utility function is known in Statistics as Scoring Rule (SR). Let $\overline{\mathbb{R}} = [-\infty, +\infty]$ denote the extended real line and the functions $H(Z) : \mathcal{P} \to \overline{\mathbb{R}}$ and $D(X, Y) : \mathcal{P} \times \mathcal{P} \to \overline{\mathbb{R}}$ be associated with any $U(Z)$.

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 when X (truly) realizes and $D(X, Y) = H(X) - S(X, Y)$ the Divergence Function between the predictive density functions of X and Y. Notice that Entropy and Divergence coincide in case of perfect evaluation of both X and Y. Finally, a hat denotes estimates while a tilde the SJB-spurred objects.

3.2 The Dynamic Scoring Structure

The object that nests all the parts of the forecasting process according to the Repeated Game reported in Supplement is defined as follows:

Definition 1 (Dynamic scoring structure)**.** We define *Dynamic Judgmental (or Scoring) Generating Structure* – or, more simply, Dynamic Scoring Structure (DSS, henceforth) – the 6-ple $\mathcal{SS} := \left\{\mathcal{Z}_t, \mathcal{F}_t, \mathcal{P}, S(\cdot,\cdot), H(\cdot), D(\cdot,\cdot)\right\}$ where Z is measured by a T-dimentional dynamic system.

Example 1. (i) A p-order autoregression with exogenous variables, ARX(p), producing density forecasts $\hat{p}(z)_{t+1}$ and and $\tilde{p}(z)_{t+1}$ using a T-dimensional dynamic

system and where FP and FU' utility corresponds, respectively to $\tilde{\mathcal{U}}(\mathcal{L}, \tilde{\pi}; z)$ and $\tilde{\mathcal{U}}(\mathcal{L}, \tilde{\psi}; z)$, where $\tilde{\pi} \in \Pi$ and $\tilde{\psi} \in \Psi$, is a DSS.

(ii) A p-order autoregression with exogenous variables, ARX(p), producing density forecasts $\hat{p}(z)_{t+1}$ and $\tilde{p}(z)_{t+1}$ using a static regression framework is a SS *but not* a DSS.

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

Strictly convexity of $\mathcal{P}$ implies to assume that $\mathcal{U}$ is strictly proper. In turn, strictly proper SRs require the use of a more general space than 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. *$\mathcal{U}(\cdot, \cdot, \cdot)$ and π are unobserved.*

Assumption 2 is challenging from a methodological point of view, because it implies that there are at least two sources of uncertainty: the utility function and judgment. Whereas the latter is known, it would still possible to recover $\hat{\mathcal{U}}$ via state-space representation of the ARX(p) and assuming a quadratic loss function, so that OLS machinery and Kalman filtering works. When also the judgment term is unknown, the classical state-space modelling is still possible, but the Kalman filter is no more the minimizer of the mean square criterion. In turn, this is due to the fact that π is treated as a non-stochastic process that can only be inferred by the repeated game assumed in the Forecasting Protocol. The next Subsection deals with this issue.

3.3 The Deformed Likelihood and q-Entropy

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

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

The behavior of (1) is displayed in Figure 1.

Consider the probability space previously defined and a continuous densify function f with parameter θ. Then, assuming a normal distribution, the Deformed Likelihood of f is:

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

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

Definition 2. Let $z_1, \ldots, z_T$ be an i.i.d. sample from $f(z_i, \theta_0)$, $\theta_0 \in \Theta$. Then the maximum Lq-estimator (MLqE) of θ_0 is

$$\hat{\theta}_T \doteq \max_{\tilde{\theta} \in \Theta} \sum_{t=1}^{T} L_q[f(z_t; \theta)], \quad q > 0. \tag{3}$$

Equation 3 is the result of the maximization of the Lq-likelihood equation, that, at the t-observation takes the form:

$$\sum_t w_t U_\theta = 0 \tag{4}$$

which is a weighted version of the likelihood equation with $U_\theta = f(z_t; \theta)'/f(z_t; \theta)$ and weights $w_t = f_t(z_t; \theta)^{1-q}$. When $q < 1$, data points with high likelihoods are assigned large weights. As q tends to 1, the MLqE coincides to standard MLE. Typically, outliers have very small weights. The estimation of (4) in a time series framework is the ultimate aim of this paper. A synopsis of the behavior of deformed

logarithm in statistical functionals is reported in Figure 2.

A considerable simplification of our treatment comes by the following:

Remark 1. If $f(\cdot,\cdot)$ is Gaussian, then the estimated mean of $\hat{\theta}$ does not depends of q.

The parameter q is a point measure of judgmental bias in the estimated model due to FP (or FU) *singularly*. That is, equation (3) measures only a non-strategic judgment. Analyzing the strategic effects of this deformation parameter requires to consider the discrepancy between U obtained (a) when event realizes (using the same $f(\cdot,\cdot)$ or one of its moments); and (b) the maximal utility obtained when no bias is assumed by both FU and FP. When this discrepancy is computable, in principle, it is possible to do comparison and hypothesis testing to verify the effectiveness of the bias, hence do claim validation on the predictive density $f_{t+1}(\cdot)$ using the empirical equivalent of (a) and (b). By Remark 1, these are the matrices of optimal (non-biased) variance and the matrices of observed, potentially biased variance. We suggest the use of the Brègman distance:

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

where $\tilde{\Lambda} : \mathbb{R}^{N\times N} \to \mathbb{R}$, with $\mathbb{R}^{N\times N}$ being the space of positive semi-definite matrices, is a scalar function three times continuously differentiable with $\tilde{\Lambda}(V_t) = \nabla\tilde{\Lambda}(V_t)$ and $\tilde{U}'(V_t) = \nabla^2\tilde{\Lambda}(V_t)$, denoting the gradient and the Hessian of $\tilde{\Lambda}$ with respect to the $K = N(N+1)/2$ unique elements of V_t and $\Lambda'(V)$ is negative semidefinite.[6] The goodness of this family of distances is confirmed by the coherence test set by Zanetti Chini (2019).

Then, the resulting Entropy function is defined as follows:

Definition 3. Let f and g be the probability density function of Y and X, respec-

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

tively. Then, the q-entropy of g with respect to f is

$$H_q(f,g) \doteq -\mathcal{E}_f L_q[g(X)], \quad q > 0 \tag{6}$$

where L_q is defined above.

Remark 2. (i) if q approaches to 1, then $H_q(\cdot,\cdot)$ nests the Shannon Entropy and $D_q(\cdot,\cdot)$ the Shannon Entropy.

(ii) The minimizer over θ of $D_r(\theta,\, \theta^0)$, where "r" denotes an alternative (biased) distribution and θ^0 is the true parameter, is the same as the minimizer $H_r(\theta_0, \theta)$ where $q = 1/r$.[7]

3.4 The State-Space Representation

We consider a univariate time series y_t observed in 1, . . . , t,. . . , n. Then the state-space form for y_t is the following system of equations

$$\begin{cases} y_t \quad = Z_t\alpha + X_t\beta + G_t\epsilon_t & \epsilon_t \sim iid(0, \sigma_\epsilon^2); \\ \alpha_{t+1} = T_t\alpha_t + W\beta + H_t\eta_t, & \eta_t \sim iid(0, \sigma_\eta^2), \end{cases}$$
$$Z = [1,\ \mathbf{0}_{m-1}]; \qquad X = [1,\ \ \mathbf{0}_{k-1}]; \qquad H = \boldsymbol{\phi}' \tag{7}$$
$$G = \begin{bmatrix} \boldsymbol{I}_{m-1} \\ \mathbf{0}_{m-1} \end{bmatrix}; \qquad T = \begin{bmatrix} \boldsymbol{\phi}' \\ G \end{bmatrix}; \qquad W = [\boldsymbol{\phi}',\ G]'$$

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

$$\alpha_0 = [\mathbf{0}_m]; \quad \beta_0 = [\mathbf{0}_k]; \quad \boldsymbol{I}_{m^2} = \boldsymbol{I} \otimes [T,\ T];$$
$$H^2 = HH'; \qquad vec(P) = \boldsymbol{I}_m^{-1} H^2. \tag{8}$$

[7]See Ferrari and Yang (2010), p 755.

Then, we invoke 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 has set to simplify the notation and treatment. The next result links the state-space representation to the DSS:

Lemma 1. (i) The DSS is never isomorphic to SS.
(ii) The DSS is isomorphic to (7).

Intuitively, the DSS and SS do not differs by the nature of the observations enclosed in the 6-ple $\left\{\mathcal{Z}_t, \mathcal{F}_t, \mathcal{P}, S(\cdot,\cdot), H(\cdot), D(\cdot,\cdot)\right\}$, but in the way in which the collection of $\mathcal{Z}_t$ is estimated – by dynamic system the former and by static regression the latter. On the contrary, the DSS and (7) may differs at most by the nature of the observations in $\mathcal{Z}_t$.

3.5 The Judgment Filter

Proposition 1. Consider the system (7). Under A1 (i)-(iv), the Judgment-filter recursive equations are:
(i) For 1,. . . , t, . . . , n,

$$v = y_t - Z\alpha_t - X\beta; \qquad F = ZPZ' + GG'; \tag{9}$$

$$C = PZ'/F; \qquad t = v/\sqrt{q_0 F^{(q_0-1)}}; \tag{10}$$

$$v_t = \alpha + C\sqrt{F}h; \qquad P_t = \alpha_t + Cq_0FC'(h/t); \qquad Q = HG'/F; \tag{11}$$

$$P_{t+1|t} = TV_tT' + HH' - (QFQ' + QFC'T' + TCFQ')(h/t); \tag{12}$$

$$\alpha_{t+1|t} = Tv_t + W\beta + Q\sqrt{(q_0F)}(h/t); \tag{13}$$

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

$$h = \begin{cases} t & if \quad (t < t^*) \\ h(t,a,b) & otherwise, \end{cases} \tag{14}$$

and

$$h(t,a,b) = \begin{cases} t; & if \quad |t| \leq a \\ \frac{a}{b-a}(b-t) & if \quad a < t \leq b, \\ \frac{a}{b-a}(b+t) & if \quad -b < t \leq t-a, \\ 0 & if \quad |t| \geq b \end{cases} \tag{15}$$

being the two-piecewise Hampel function.

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

$$L = log(F) + log(q_0); \qquad S = v^2/F; \tag{16}$$

$$y_f = Zv_t + GG'\sqrt{q_0 F}h/(q_0 F); \qquad h_t = h/t; \tag{17}$$

$$\alpha_{t+1} = a; \qquad \Sigma_{t+1} = diag(P) \tag{18}$$

Proof. See Appendix □

Corollary 1. *From the above recursions we get the estimated weights:*

$$w_t = 1/(L)^{q_0}\mathcal{G}_t \tag{19}$$

and the averaged measures:

$$\mathcal{L} = -0.5[T\log(2\pi) + L + S]; \qquad \mathcal{L}c = -0.5[T(\log(2\pi\mathcal{S}) + 1] + L; \tag{20}$$

$$\mathcal{L}_q = 0.5((q_0 T\log(2\pi) + (L+S)^{q_0}); \qquad \hat{\mathcal{L}}_q = 0.5((\hat{q}T\log(2\pi) + (L+S)^{\hat{q}}). \tag{21}$$

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

Equation (19) is the core result of this paper. In facts, it defines the weight of deformation parameter q associated to each period via the gradient vector $\mathcal{G}_t$.

Finally, the following result ensures that JF and KF are strictly related:

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

Proof. Let DSS_1 the DSS under Lq-transform and DSS_2 an equivalent without Lq-transformation and the operator $\mathcal{T}$ denotes the state-space system (7). It suffices to note that $\mathcal{T}$ is the same in either DSS_1 and DSS_2, since DSS_2 is a DSS_1 with q=1. See Remark 1. □

4 Application

This section applies the DSS-JF to real macroeconomic data. Subsection 4.1 describes the data; subsection 4.2 reports the results and 4.3 provides a discussion on them.

4.1 Data

The SPF-FED is the oldest dataset on the professional forecasting activity on a macroeconomic dataset, being the survey observations started in 1968. The survey was originally administrated jointly by NBER and the American Statistical Association and, since the second quarter of 1990, these two institution has been replaced by the Real Time Data Research Center of the Federal Reserve of Philadelphia. In more that 50 years, the SPF-FED has changed considerably, increasing its amplitude (from 10 variables to forecast in the first survey in 1968, to several dozens in the current release)[8] and its complexity (with increasing number and technicality of therein items); see the SPF website[9] and Croushore et al. (2019).

[8] We refer to the Release 2020 Q3.

[9] https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

In particular, our research effort is stimulated by the fact that SPF-FED respondents are given a considerably short time to answer to the survey.[10] Albeit assuming that survey has been conducted by the best human and computational capital in the market, so that either publicity-seeking behavior (Laster et al., 1999) either noise-information (Orphanides, 2003) among the most popular arguments for criticism of the survey may be annihilated, the use of non-sample estimates in forecasters output cannot be neglected completely. This, in addition to the change of ownership and constant engineering of the surveys make us to suspect a change in the amount of judgment in the sample span.

This paper focus on Real GDP growth as case study variable due to its universal use for monitoring and addressing the economic policy by deputed institutions. A set of 4 additional variables are also included to verify the behavior of the DSS-JF with exogenous variables. These, representative of different economic sectors, are: production, consumption, money velocity and house prices; see Table 1. We remark that there is not an assessment of the monetary policy rather other issue, while this is only an experiment of a new methodology. We apply our JF over 10 systems of variables where all the combinations are considered. All the real data apart the SPF forecasts has been downloaded by FRED in quarterly frequency and, whenever possible, seasonally adjusted.

4.2 Results

The estimates of the aggregated quote of unobserved SJB and the main (deformed) functionals and variance parameters are reported in Tables 2 and 3. When the DSS-JF is applied on 'one-side' real data of RGDP (that is, the data without FU's outputs as exogenous variable) the estimated aggregate judgment is zero, and thus, $q = 1$ (Table 2 first column of top panel), so that the DSS-JF coincides with a standard State-Space form of an AR model where the JF coincides with Marczak et al. (2018)

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

filter. Figure 3 plots the estimates of the corresponding dynamic system. In this peculiar case we did two trials, each one corresponding to a different initial value of q. In both the cases the estimated parameter is always 1 (that is, zero effect in judgment), thus ensuring a high credibility to our method.

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, being $\hat{q} = 0.32$ (that is more than two third of the estimates are due to judgment) for nowcasts and arriving to zero for long-run horizon forecasts, meaning that these last are purely judgmental.

Such an extreme variation among the estimates based on historical data and forecasts may be easy explained by an omitted variable bias. Thus, we replicate our estimation exercise with all the systems defined in last panel of Table 1. The results are reported in the second panel of the same Table: noticeably, the estimated quote of judgment is always minimal being the $\hat{q}$ 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 quite more realistic scenario, albeit the ARX-DSS of this peculiar case is based only on historical data and not on forecasts. This means that whenever $\hat{q} \neq 1$ the DSS 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 judgmental bias.

The estimates resulting from the ARX-DSS based on one-quarter-ahead forecasts (fourth panel) are instead characterized by a non negligible amount of SJB, being estimated q never less more than 0.80, this time without differences among the different models. Similar results, with possibly stronger evidence of judgmental bias holds in higher horizons, reported in Table 3. In general, Models M5 and M9 seem the most judgmental-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 the fact that all the deformation parameters are above the same thresholds.

The previously stated results demonstrate that the forecasts of U.S. RGDP are characterized by, approximately, 20% of bias due to strategic judgment. When this bias is more evident? Figures 4 – 6 answer to this question. The outputs of the application of DSS-JF on univariate time series, plotted in Figure 4, panel a, is coincident with Figure 3, panels (c) and (d) if considering only historical data, while other univariate time series corresponding to SPF nowcasts and forecasts varies considerably in a similar fashion of Table 2. Things change when looking 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 tends to appear only in correspondence of crisis. The one-quarter-ahead forecasts are instead characterized by a weight function generally near-zero with important jumps, the frequency of which varies considerably according to the model.

On a different side, the two/three/four-quarters-ahead forecasts – reported in Figures 5 and 6 – are characterized by weighting function lying in 1 with a small number of negative peaks, generally in the first half of the sample with the only exceptions of the Pandemics, where the evidence is in favor of the judgment is considerable. However, the timing of these big changes in weighting function is not the one that 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, to arrive to zero in 2005-2006 and (only in a few models for one quarter-ahead forecasts) in 2012-13. Interestingly, the official NBER recessions date for the Great Recession are never involved.

4.3 Discussion

Several considerations can be made from the above empirical investigation: first, there is a dynamics in the aggregate quote of SJB 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 the econometric and computing techniques, that gradually made the use of statistical tools more affordable for the economists 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).

Second, the non coinciding dynamics of the switches, in the weighting function, of the majority of models with NBER recession dates implies that SPF does not changes their attitudes mechanically or according to a simple algorithm like "*use-judgment-when-recession-arrives*", but several different, more complex reasoning schemes are beyond their output. This seems consistent with the Coibion et al. (2018)'s claim for a novel rethinking of the expectation formation. In this sense, the mechanics of endogenous formation of mis-judgment that can be inferred by Ilut and Valchev (2020) is confirmed by our evidence.

This finding opens a question on the effectiveness of data revisions that FED makes periodically. In facts, if a certain amount of judgment bias is proved, the role of this bias 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 implies that these last should coincide, or be in a neighborhood, with the degree of judgment found in our analysis. Instead, the revisions certified by Bureau of Economic Analysis (BEA) in the period 1993–2019 seems to reject this hypothesis: the role of revisions has been considerably small, only of among 0.5 and 1.2 percentage points of the estimates (in average), while a simple graphical inspection of Figure 3 and Tables 1 – 2 suggests the bias is higher[11]. A future development of our DSS

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

methodology that takes into accounts also the effect of FED's staff revisions is highly recommendable.

We remark that the our result that the SJB is not negligible – in general, and, in particular, in some periods – should not be interpreted necessarily as a claim that FED staff does not uses efficiently all the available amount of information. The efficiency of the FED administration has been empirically tested, among many others, by Messina et al. (2015), who support the noise-information hypothesis. Instead, we interpret this finding as the survey respondents diversify their efforts among computational econometric mechanics and human deliberation. This interpretation is coherent with the finding by Casey (2020) that 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. However, this interpretation should by combined with the further recent results by the same author (Casey, 2021) suggesting a general over-confidence of professional forecasters. This make us to question the nature of the link among overconfidence, uncertainty and judgment. Further theoretical research is necessary to this aim.

Finally, we aware about the limits of our methodology. Despite the DSS is a very general framework that applies a (time series) regression framework on data coming from two different agents with (possibly) different utility functions. In this application we assumed only a single, representative forecaster that materializes as average of multiple forecasts. The current DSS form does not allow to discriminate the contribution of each single individual, unless repeating the same DSS on individual data. Moreover, the Deformed Likelihood estimator, due to its logarithmic structure, may be not the best option in case of repeated large outliers and some extreme-value method may be preferable; see, among others, Burridge and Taylor (2006). Finally, a recent strand of literature that studies the predictive power of ultra-high-dimensional micro-data owned by web firms finds that SPF-FED forecasts are not superior to the ones computed by using the former; see, among other,

D'Amuri and Marcucci (2017). Our DSS approach is still unset for this kind of analysis. A better integration of the econometric methods for large data and our DSS framework may be useful to better explain the forecaster disagreement and, in prospective, the different predictive power difference among these two approaches.

5 Conclusions

The macroeconomic forecasts for the U.S. economy are characterized by a mixture of data and non-data-driven (or judgmental) estimates. The exact partition of these elements has been estimated for the first time in econometric literature. This has been possible by combining several statistical methods, and namely, the Deformed Likelihood estimation, (robust) signal extraction and linking the resulting statistical model to a repeated game and and a set of decision-based rules. The corresponding DSS-JF allows econometricians to extrapolate the SJB from a time series of forecaster's output using a set of exogenous variables. It corresponds to classical Autoregressive State-Space Model to be estimated via Robust Kalman Filter if forecasters agents are considered as an homogeneous aggregate.

Our simulation experiments reveal that the DSS does not tend to over-evaluate the judgmental quote in forecasting activity. We then applied the new method to U.S. forecasts of Real GDP. The evidence supports the hypothesis that FED-SPF has a non negligible amount of judgment and this is distributed non-uniformly during the span of the sample. This confirms the recent claims by several authors that the rational-expectation mainstream framework, despite the recent refinements, should be severely modified or substituted by a more realistic hypothesis.

The methodology here proposed is at the beginning step of development. More research is required to understand the motivations of such a judgmental dynamics and the potential capability of DSS analysis in ultra-high-dimensional forecasting environments.

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.

Al-Najjar, N.and Pomatto, L., Sandroni, A., 2014. Claim validation. American Economic Review 104, 3725–36.

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.

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.

Burridge, P., Taylor, R., 2006. Additive outlier detection via extreme-value theory. Journal of Time Series Analysis 27 (5), 685–701.

Canova, F., 2011. Methods for applied macroeconomic research. Princeton university press.

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.

Casey, E., 2021. Are professional forecasters overconfident? International Journal of Forecasting 37 (2), 716–732.

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., Kamdar, R., 2018. The formation of expectations, inflation, and the phillips curve. Journal of Economic Literature 56 (4), 1447–91.

Conlisk, J., 1996. Why bounded rationality? Journal of economic literature 34 (2), 669–700.

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

Dawid, A., 1982. The well-calibrated bayesian. Journal of the American Statistical Association 77 (379), 605–610.

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

De Jong, P., 1991. The diffuse kalman filter. The Annals of Statistics, 1073–1083.

Dekel, E., Feinberg, Y., 2006. Non-bayesian testing of a stochastic prediction. The Review of Economic Studies 73 (4), 893–906.

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.

D'Amuri, F., Marcucci, J., 2017. The predictive power of google searches in forecasting us unemployment. International Journal of Forecasting 33 (4), 801–816.

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

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

Giannone, D., Lenza, M., Primiceri, G., 2021. Economic predictions with big data: The illusion of sparsity. Econometrica In press.

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

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., 2020. Economic agents as imperfect problem solvers. Tech. rep., National Bureau of Economic Research.

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.

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.

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.

Messina, J., Sinclair, T., Stekler, H., 2015. What can we learn from revisions to the greenbook forecasts? Journal of Macroeconomics 45, 54–62.

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

Olszewski, W., 2015. Calibration and expert testing. In: Handbook of Game Theory with Economic Applications. Vol. 4. Elsevier, pp. 949–984.

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

Ottaviani, M., Sørensen, P. N., 2006. The strategy of professional forecasting. Journal of Financial Economics 81 (2), 441–466.

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.

Pomatto, L., Al-Najjar, N., Sandroni, A., et al., 2014. Merging and testing opinions. Annals of Statistics 42, 1003–1028.

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.

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.

Varian, H. R., 2014. Big data: New tricks for econometrics. Journal of Economic Perspectives 28 (2), 3–28.

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.

Zanetti Chini, E., 2019. Strategic judgment: its game-theoretic foundations, its econometric elicitation.

A Appendix

Proof of Proposition 1

(i) CASE 1: q=1. In this case, the Lq-Likelihood function coincides with the standard Likelihood and the Judgment Filter coincides with the classical Kalman Filter (or its robust version). 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 MMSLE of α and β. Thus, we need only to justify the introduction of (a) $t = v/\sqrt{q_0 F^{q_0-1}}$ in eq. (10) and (13); (b) q_0 in the second addend of (11).

Let start from (b) According to the Ferrari and Yang (2010), the surrogate parameter for the normal equation in a Gaussian regression is $\theta_t = (\mu', \sqrt{q} vech(\mathbf{\Sigma}))'$. We have 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, it suffices to notice that there is no need of half-vecotrizing P. The same argument holds for equation (13).

To prove (b), let remark that, in exponential family, $\theta^* = \theta_0/q$, where $\theta = [\mu, \sigma^2]$. Since μ is not influenced by q, it suffices to consider σ^2/q, where σ^2 is known. Thus, the only unknown variable is $1/q$. According to the Ferrari and Yang (2010) definition of normal equation, the optimal solution is the first derivative of $\sqrt{F^q}$

(ii) Direct consequence from (i).

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 4.

Table 2: Application of DSS on a real data

Parameter	AR-DSS					
	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-DSS 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-DSS 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-DSS 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

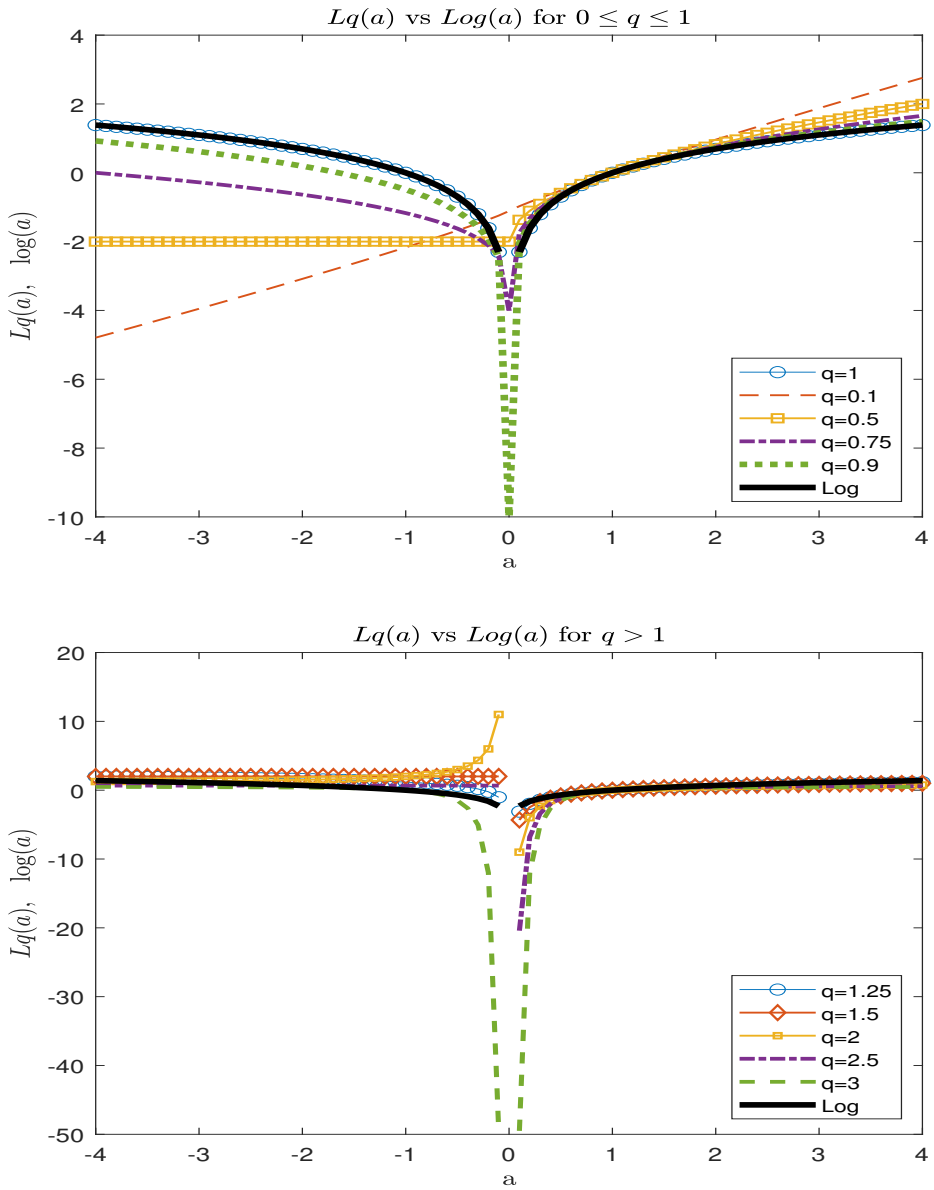
NOTE: This table reports the estimates of the DSS-JF exposed in Section 3. The top panel concerns about pure Autoregressive case (that is, the application of the DSS-JF to the single time series of real data, nowcasts and forecasts by SPF; the further three panels concerns the systems with exogenous variables described in Table **??**. In these last, the dependent variable is, respectively, the historical value, the nowcast and 1-quarter-ahead forecast of RGDP.

Table 3: (Continue...)

Parameter	ARX-DSS 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-DSS 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-DSS 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

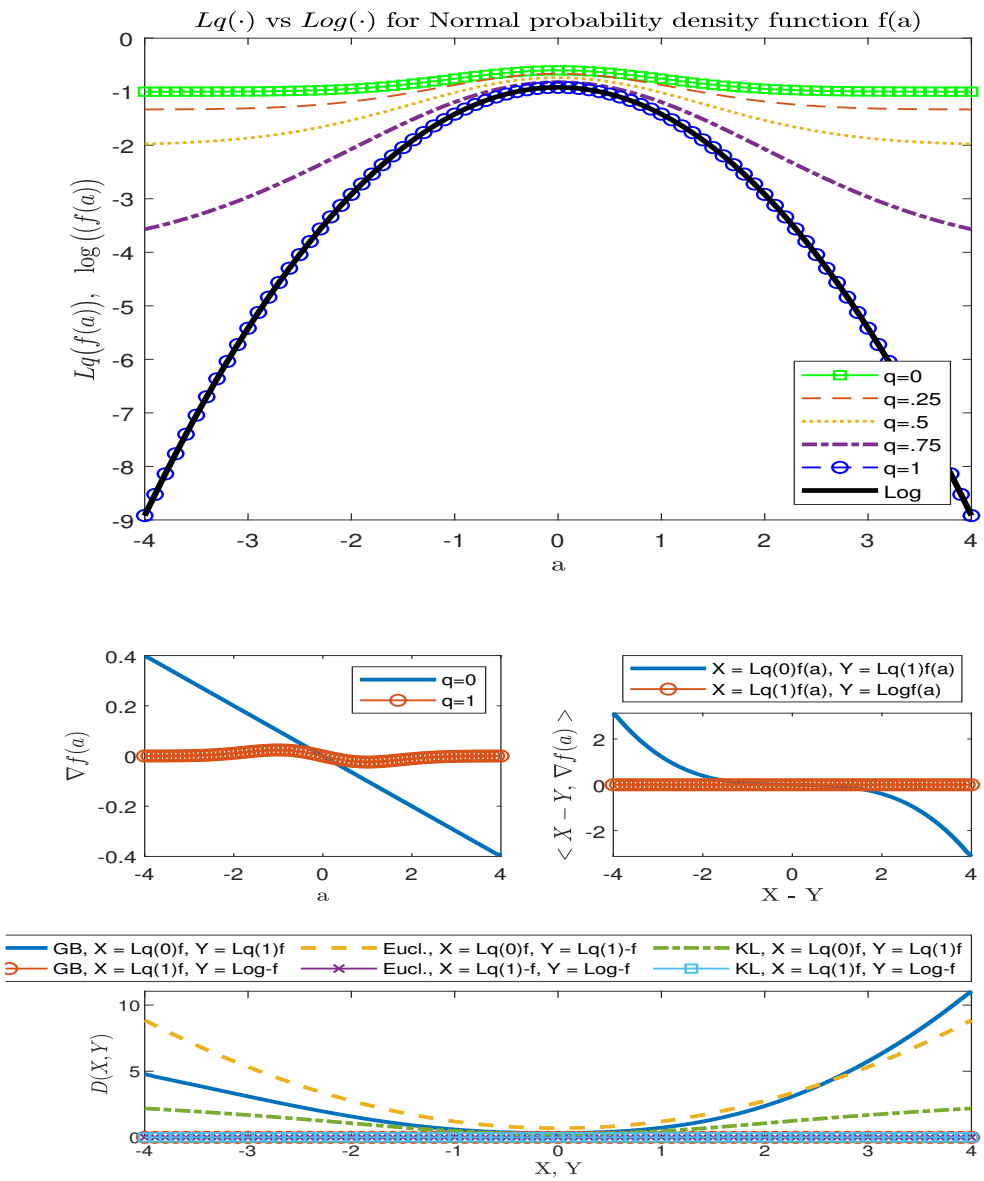
NOTE: This table continues to report the estimates of the DSS-JF exposed in Section 3. In this case, the same systems with exogenous variables described in Table. **??** are analyzed, but the dependent variable is a forecast with higher horizon.

Figure 1: The deformed logarithm function



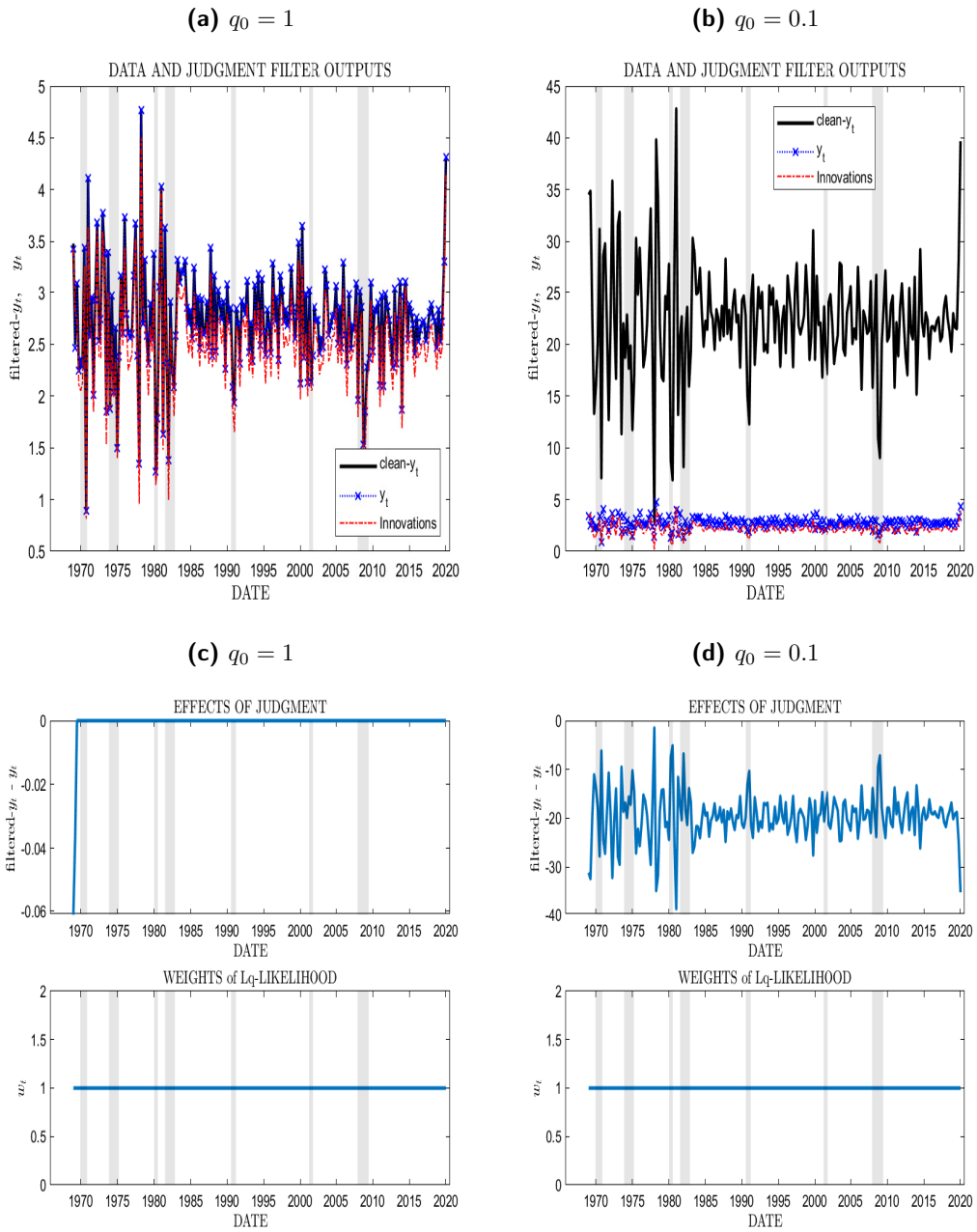
NOTE: This figure displays the Lq-function applied to the numerical sequence $a = [-4; 4]$ for different values of q and compares it with the natural logarithm function.

Figure 2: Functional analysis of the Deformed Logarithm



NOTE: This figure displays the behavior of several functionals of the Deformed Logarithm. Left 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. Right 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, either in the case of $X = Lq(f)$, with q=0 and $Y = Lq(f)$ with q=1 (blue), either 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).

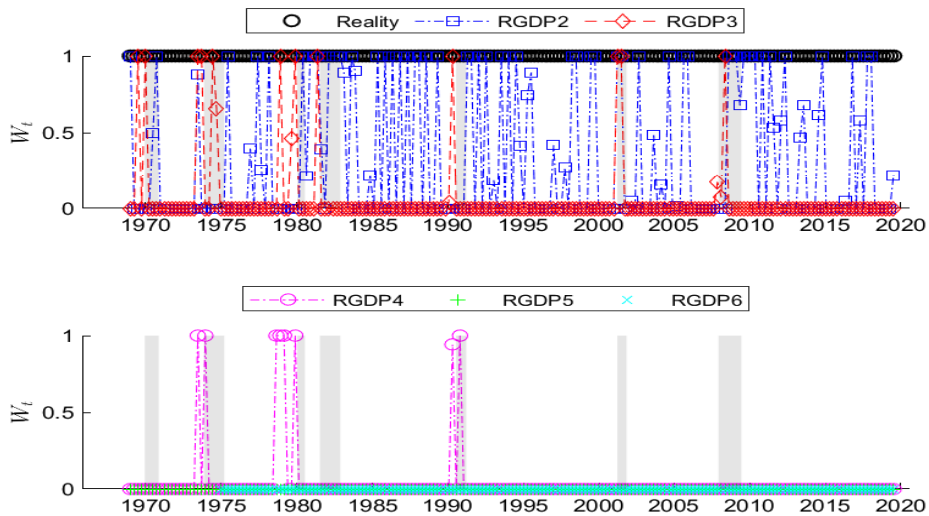
Figure 3: Estimation of unobserved judgment on U.S. Real GDP historical data



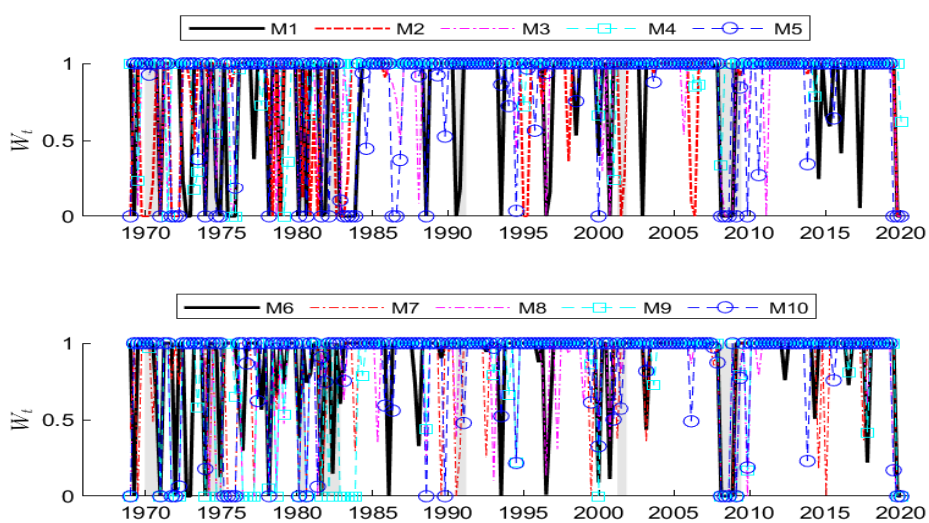
NOTE: This figures plots the results of the application of the Deformed Kalman Filter on the data of U.S. Real GDP by Federal Reserve Bank of St. Louis. Left panels deal with $q_0 = 1$ assumed in the DSS, while right panels display the results of a DSS with $q = 0.1$.

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

(a) Historical data



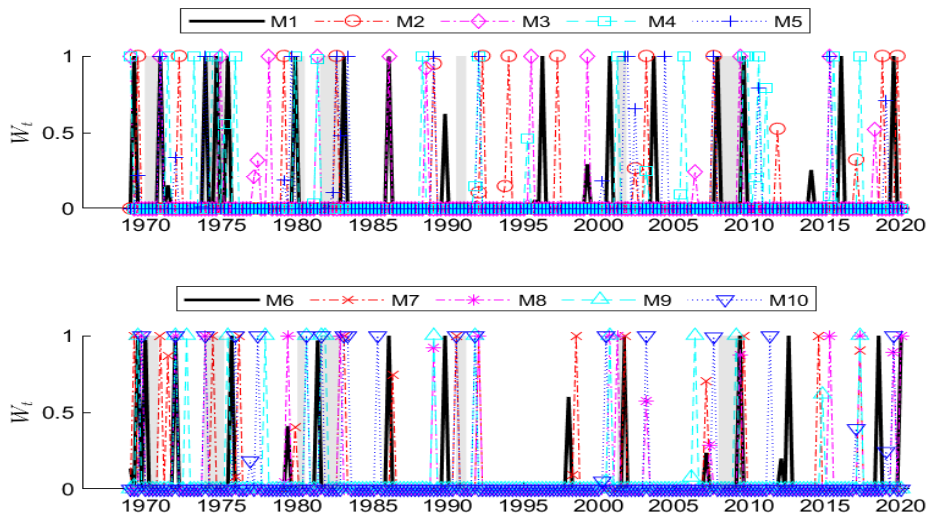
(b) Nowcasts



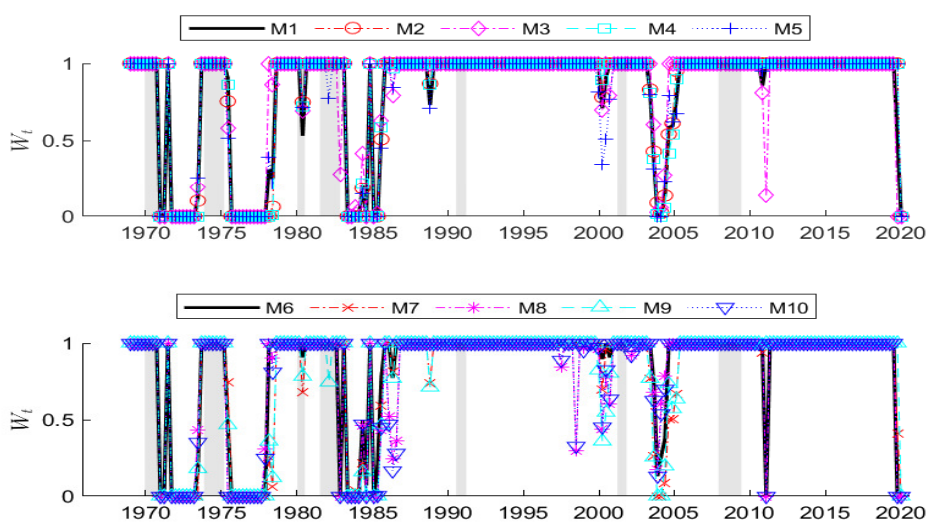
NOTE: This figures plots the results of the application of the DSS on U.S. Real GDP forecast by SPF using several data and model specifications.

Figure 5: Estimation of unobserved judgment on U.S. Real GDP forecasts using several DSS specifications

(a) 1-quarter-ahead



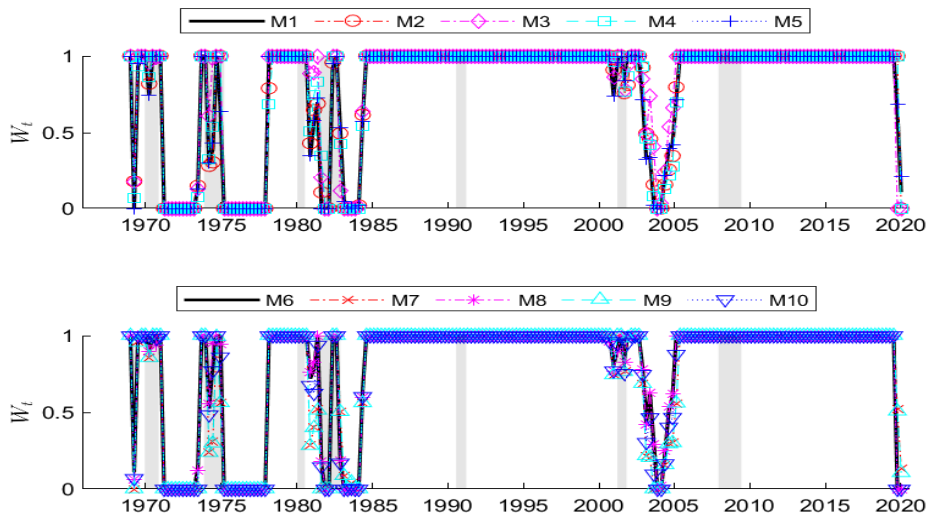
(b) 2-quarter-ahead



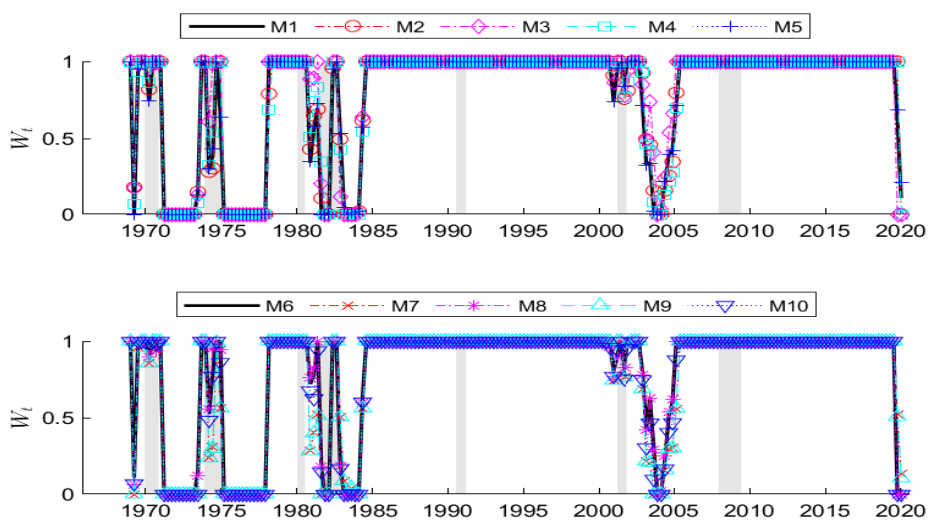
NOTE: This figures plots the results of the application of the DSS on U.S. Real GDP forecast by SPF using several data and model specifications.

Figure 6: Estimation of unobserved judgment on U.S. Real GDP forecasts using several DSS specifications

(a) 3-quarter-ahead



(b) 1-year-ahead



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

SUPPLEMENT
to
"Measuring Unobserved Strategic Judgment"

Emilio Zanetti Chini

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

October 2021

1 Introduction

This Supplement includes further results that corroborate the treatment of Main Document. Namely, Section 2 defines the Repeated Game that is assumed in the Forecasting Process; Section 3 includes some basic results which knowledge is assumed in Main Text; finally, Section 4 reports the Monte-Carlo simulation of the DSS.

2 The Repeated Forecasting Game

We assume that the probabilistic forecast of an economic event is the output of a repeated game with three players: the FP; the FU who has capital K to preserve; and Reality. The FU suspects the FP's quotations are biased and, eventually, cooperates with Reality; however, no matter how the FU plays, Reality acts as though the FU does not win the game. This rule, called "*Cournot's Principle*", is necessary to

avoid that the game is unbalanced in favor of FU. These players act according to the *Forecasting Protocol* here defined:

Definition 1 (Forecasting Protocol)**.** *For* $i = 1, \ldots, n$,

1. $K_0 := 1$;
2. FU announces a bounded function $S_i : \mathbb{R} \to \mathbb{R}$;
3. FP announces her (potentially biased) quotation $\tilde{p}_i(Z) \in \mathbb{R}$ where $\tilde{p}_i(Z) = \hat{p}_i(Z) + \pi_i$
4. Reality announces a draw from $P_i(Z) \in \mathbb{R}$;
5. $K_i = K_0 + D\left(Y_i, X_i\right)$,

FU must choose S_i so his capital remains non-negative ($K_i \geq 0$) no matter what values the FP and Reality announce for $\hat{p}_i(X)$ and $P(Y)$. The winner is the FU if $K_1 >> K_0$. Otherwise, the FP wins.

The game illustrated here is a re-proposition of the "Forecasting sub-game" by Vovk and Shafer (2005, p. 754) and a generalized version of the one used in Zanetti Chini (2019).

The Step 2 of the Protocol can be interpreted as one of Patton (2019)'s main conclusion: utility-based objects like the forecast rankings are generally sensitive to the choice of a proper SR, and, as a consequence, the FPs should be told *ex-ante* what utility functions will be used to evaluate their quotations[1].

The Step 5 of the Protocol is a test for the null hypothesis of forecasting coherence in terms of the FU's utility. The form in which the test is written implies that the FP's reward cannot be augmented after his quotation. This coherence test is essentially based on the $D-$function defined in previous Subsection. In principle, the assumption that Reality can cooperate with the FU implies that, when the game is repeated n times, the sequences of outcomes S_n Y_n, X_n do not necessarily coincide with realizations of a stochastic process. As a consequence, classical hypothesis

testing and inference is ineffective and should be substituted by another type of inference who explicitly accounts for strategic behavior, see Olszewski (2015) for a theoretical discussion of this problem. Nevertheless, Shafer and Vovk (2005, Chapter 8.1) ensures that the Cournot's Principle allows both of them to be used. However, in this paper we assume that π exists (so that data are affected by SJB), coherently with the evidence by Zanetti Chini (2019), so that the test is not discussed nor applied for economy of space.

3 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. Any contamination 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).

4 Simulation

This section investigates the empirical properties of the DSS-JF in a MonteCarlo simulation exercise. Namely, Sub-section 4.1 describes the Data Generating Process (DGP) adopted; Sub-section 4.2 reports the results; finally their relevance is discussed in Sub-section 4.3.

4.1 The Data Generating Process

We consider two different DGPs:

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

and

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

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

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

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 in 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 a special attention to the role played by who look at the analysis: in this case, (3) represents what FP observes and analyses (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 a time series of FU's final announcement, the whole DSS may be interpreted as a full dynamic system where the FP's output (in this case, x_t) is an input that co-exists with reality, and their discrepancies is the basis for an ex-post assessment of forecast user via utility function. This last is incorporated in estimation step via the Lq-Likelihood (see previous Sub-section). Hence, q_0 represents also a sort of a-priori of FU with respect to which FP adjusts her 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.

4.2 Results

A draw of the simulated AR(2) process in equation (3) without outlier is displayed in Figure 1, while the same process with additive outliers can be verified in Figure 2. 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 3. 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 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 Main Text and the estimated distortion parameter computed from the complete MonteCarlo exercise are reported in Tables 1 – 2. 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, no matter of 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 50%, approximately.

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 2, where $\mathcal{L}_{q0} = 3,875$ and $\mathcal{L}_{\hat{q}} = 9,357$ (more than the double). Such an inflation effect is generally more pronounced in the case of high initial q_0 and lower for low initial value of the same parameter.

Third, the predicted error variance (PEV) as well 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).

4.3 Discussion

Despite the simplicity of the DGP assumed in our simulation exercise we can do some important conclusions about the introduction of a SJB in a time series process: first, any judgmental contamination 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 the line with the theoretical fundaments by Ilut and Valchev (2020) on the dynamics of the cost of the deliberation. In facts, according to these authors, the true policy function corresponding used by FU in her evaluation is unknown. This reflects in uncertainty on the true SR, which can only be inferred by FP by bayesian methods to update her beliefs. Thus, FP gradually accumulates information about the optimal quotation as function of the underlining state. Such an accumulation is responsible of the propagation of the judgment in all the 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 a possibility to have an estimate of the amount of SJB that may diverge from q_0 in any new recursion.

Finally, this simulation exercise enlightens on the connections among our DSS

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

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 FP 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 2), where the only discrepancies from 1 in the weighting function coincide with the timing of outliers. Thus, under normality and perfect specification, the FU and FP learn and update their beliefs immediately. On the other side, when the FU misjudges the FP's initial amount of judgment, the resulting signal generates a noise that adds to the forecasting process forcing the FU to deliberate frequently 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 1 and 2, 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.

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., 2020. Economic agents as imperfect problem solvers. Tech. rep., National Bureau of Economic Research.

Olszewski, W., 2015. Calibration and expert testing. In: Handbook of Game Theory with Economic Applications. Vol. 4. Elsevier, pp. 949–984.

Patton, A., 2019. Comparing possibly misspecified forecast. Journal of Business & Economic Statistics 38 (4), 796–809.

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

Shafer, G., Vovk, V., 2005. Probability and finance: it's only a game! Vol. 491. John Wiley & Sons.

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.

Zanetti Chini, E., 2019. Strategic judgment: its game-theoretic foundations, its econometric elicitation.

Tables and Graphs

Table 1: Simulation of the DSS

# 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.

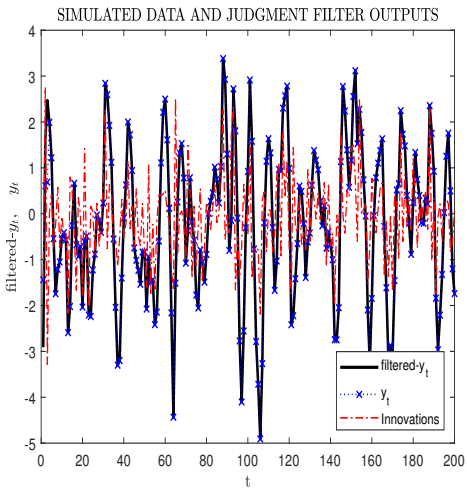
Table 2: Simulation of the DSS with innovation outliers

# 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	-	-

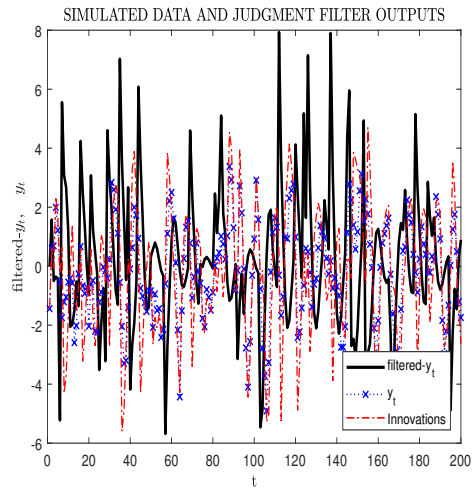
NOTE: This table reports the results of a Monte Carlo simulation of the AR(2) in equation (4) for 5,000 replications.

Figure 1: Application of the Deformed Kalman Filter on simulated data.

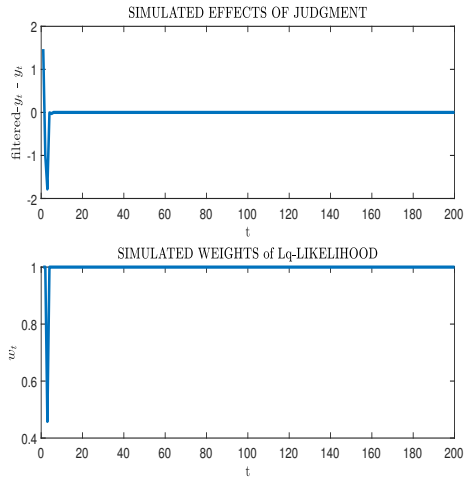
(a) The simulated process with $q_0 = 1$



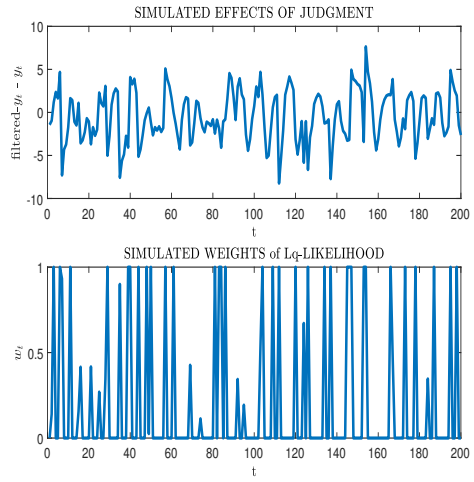
(b) The simulated process with $q_0 = 0.1$



(c) Filter components when $q_0 = 1$



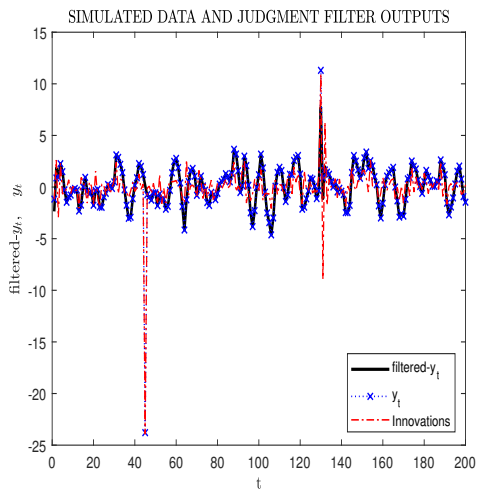
(d) Filter components when $q_0 = 0.1$



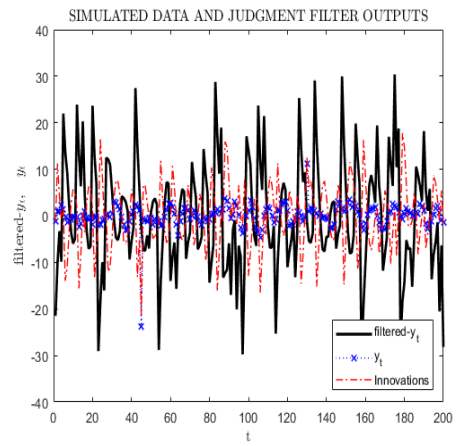
NOTE: This figure plots the result of an judgment filtering exercise using simulated AR(2) process described in (3) without innovation outliers. The upper panels display the original data, the "clean" process and the innovations. The lower panels display the output of the spread clean process vs the 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 2: Application of the Deformed Kalman Filter on simulated data with innovation outlier.

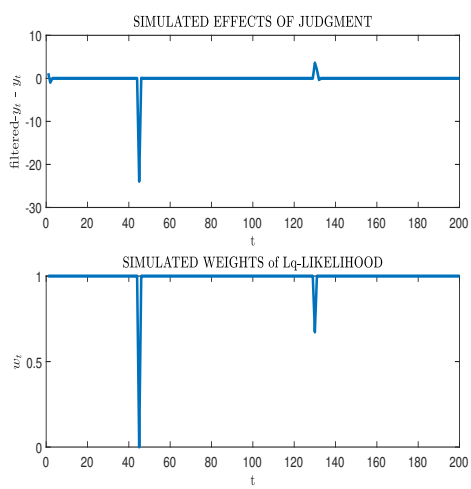
(a) The simulated process with $q_0 = 1$



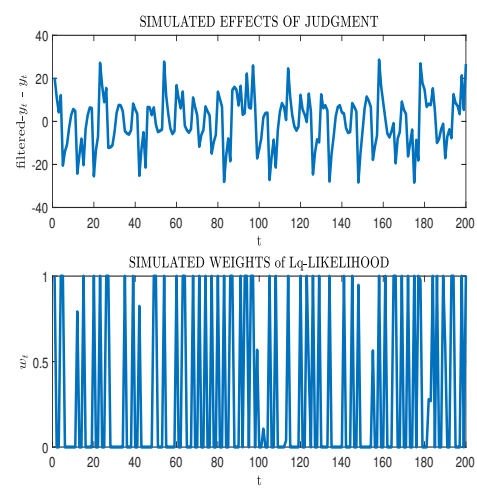
(b) The simulated process with $q_0 = 0.1$



(c) Filter components when $q_0 = 1$

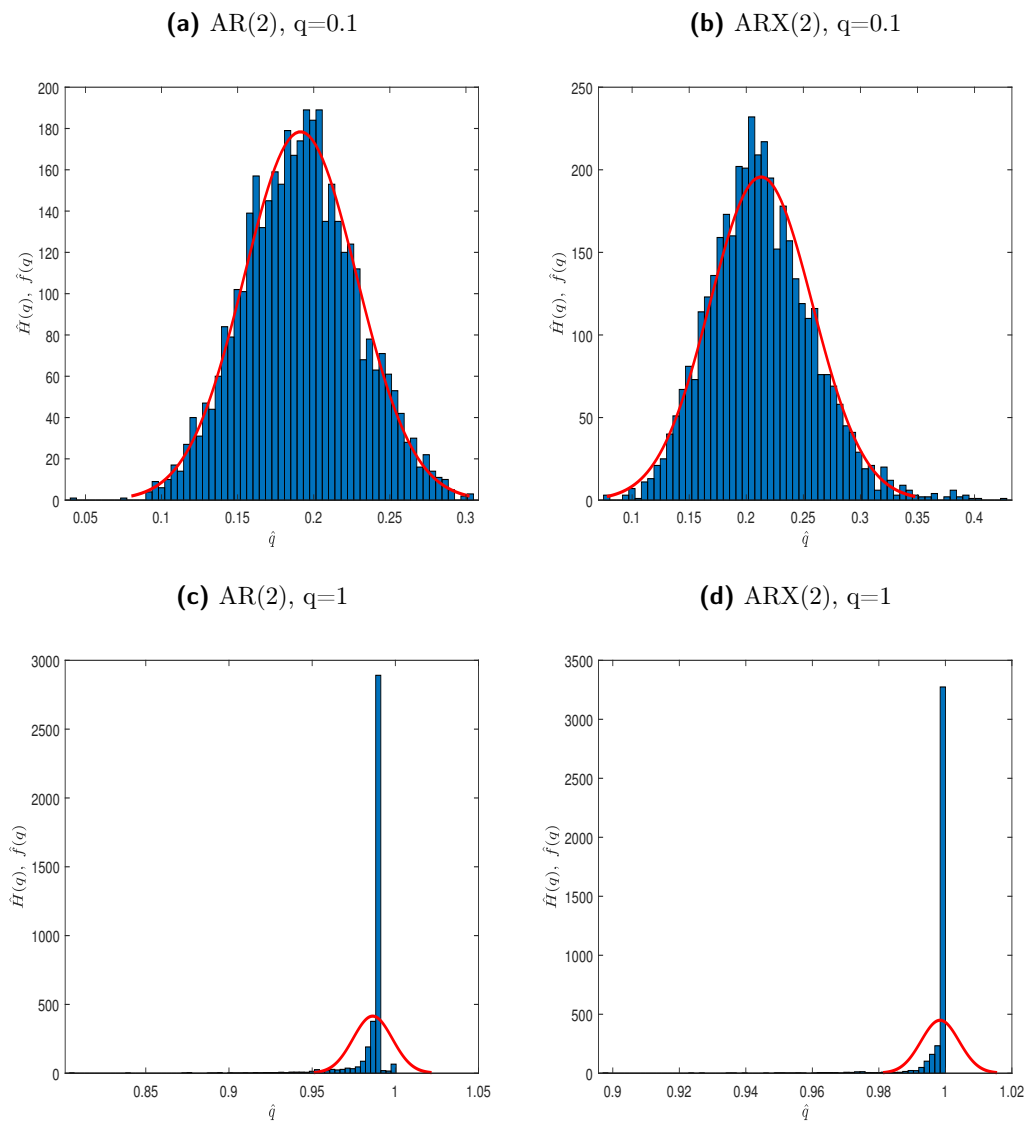


(d) Filter components when $q_0 = 0.1$



NOTE: This figure plots the result of an 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 3: 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 4. 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.