Robust Forecasting

by Timothy Christensen, Hyungsik Roger Moon, Frank Schorfheide

11th European Central Bank Conference on Forecasting Techniques: Macroeconomic Forecasting in Abnormal Times

Discussion: Tatevik Sekhposyan

Background

→ Vector autoregressions (VAR) are common forecasting tools in macro/finance.

→ VAR (in a companion form)

$$y_t = \Phi y_{t-1} + u_t, u_t \text{ is } i.i.dN(0, \Sigma).$$

→ Predictive density: $y_{t+1|t} \sim N(\Phi y_t, \Sigma)$.

→ The sample counterparts, $\widehat{\Phi}$ and $\widehat{\Sigma}$, are identified from the data.

→ Growing literature concerned with set-identification-robust inference in the structural VARs (Giacomini and Kitagawa, 2018, etc.), but that does not effect the *unconditional predictive distributions*.
- □ $\Sigma = B_0^{-1} \Sigma_w B_0^{-1'}$

? Might not be the case for *conditional distributions*.

Overview

→ Set-identification, however, matters in some context, where different parameters in the identified set lead to different forecasts, some more accurate than the others.

→ Discrete choice models for panel data.

→ Vast literature deals with identification issues in the context of panel data models, but these models are not extensively used for forecasting.

→ This paper provides a valuable addition to the literature to think of the relevance of set-identification in the context forecasting, as well as the use of panel data methods.

Ingredients

1. Situation where there are multiple parameter values consistent with the observed distribution, yet some can generate more accurate forecasts than others (binary discrete choice is the prime example).

2. Operate with a decision-based loss function – "statistical decisions are like any other decision in that they should be driven by the goals and preferences of the particular decision maker."

   This paper has both aspects and they are equally important for calculations, derivations, proposed algorithms, etc.

Partial Identification

→ Panel data allows identification of models that would not be identified using single-outcome data.

→ Yet, in many situations you get set-identification as opposed to point identification and it can matter for forecasting.

→ Here the time-dimension is negligible relative to the size of the cross-section. Otherwise, it becomes a small-sample time-series problem.

→ Most importantly, the identification depends on the model setup, parametric choices, covariate choices (discrete versus continuous).

Partial Identification

→ Binary choice model (Arellano and Bonhomme, 2011, p. 405):

$$y_{it} = \mathbf{1}\{x'_{it}\phi + \alpha_i \geq \nu_{it}\}$$

→ y_{it} is discrete, binary (0 or 1), example labor force participation, retirement

→ x_{it} exogeneous observable

→ α_i individual-specific, unobserved variable

→ ν_{it} is i.i.d. draws from a known dist. F, independent of x_{it} and α_i

→ Average likelihood of an individual observation

$$Pr(y_i|x_i;\phi) = \int \prod_{t=1}^{T} \underbrace{F(x'_{it}\phi + \alpha_i)^{y_{it}}[1 - F(x'_{it}\phi + \alpha_i)]^{1-y_{it}}}_{\text{no direct counterpart in the data}} f_{a|x}(\alpha_i|x_i)d\alpha_i$$

Partial Identification

→ Identification issue comes from the individual effects being unobserved to the econometrician.

→ Is there a unique value of ϕ such that the model implied probabilities would be consistent with observed probabilities for some conditional distribution of individual effects $f_{a|x}$?

$$Pr(y_i|x_i) = \int \prod_{t=1}^{T} F(x_{it}'\phi + \alpha_i)^{y_{it}}[1 - F(x_{it}'\phi + \alpha_i)]^{1-y_{it}} f_{a|x}(\alpha_i|x_i)d\alpha_i$$

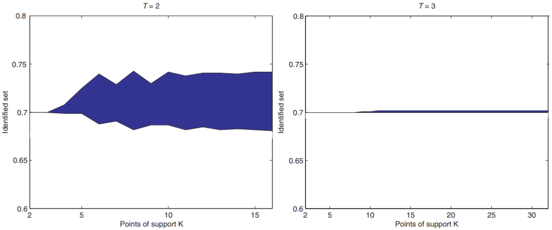
→ Chamberlain (2010) finds that with $T = 2$, and exogeneous covariates with bounded support, ϕ is point identified only when you have a logistic distribution.

→ In general we have set-identification, where the identified set Θ_0 consists of all values $\phi, f_{a|x}$, which make the model implied probabilities consistent with the probabilities observed in the data.

Partial Identification: Example

→ $y_{it} = \mathbf{1}\{\phi(t-1) + \alpha_i \geq \nu_{it}\}$, $\phi = 0.7$, probit model

→ Uniform points of support for α_i, normal cond. distr.



→ Excess support in outcomes relative to individual effects may lead to point identification.

→ Some parameterizations that are not-distinguishable based on T observations, might become identified based on $T+1$.

Objective

→ Set identification in the parameters, maps to uncertainty about forecast distribution.

→ Two types of forecasts

1. Robust forecasts — accounts for model uncertainty, $\theta \in \Theta_0$, Θ_0 is known.
2. Efficient robust forecasts — also accounts for parameter uncertainty, since Θ_0 needs to be estimated.

? Very much related to the notions of uncertainty (Knightian uncertainty, uncertainty about the distribution) and risk.

Robust Forecasts

→ The risk of $d \in [0, 1]$ under the forecast distribution P_θ

$$E_\theta[\ell(Y, d)] = \ell(0, d)P_\theta(Y = 0) + \ell(1, d)P_\theta(Y = 1)$$

→ θ-optimal forecast, minimizes average risk, d^*_θ, solves

$$E_\theta[\ell(Y, d^*_\theta)] = \inf_{d \in D} E_\theta[\ell(Y, d)]$$

→ Quadratic loss: $\ell_q(y, d) = (y - d)^2$

→ θ-optimal forecast

$$d^*_\theta = E_\theta(Y) = P_\theta(Y = 1)$$

Robust Forecasts

→ Define extreme forecast probabilities

$$p_L := \inf_{\theta \in \Theta_0} P_\theta(Y = 1) \quad p_U := \sup_{\theta \in \Theta_0} P_\theta(Y = 1)$$

→ A linear programming algorithm to obtain these probabilities when outcome probabilities are related to individual effect probabilities in a linear manner and the distribution of latter has discrete support.

→ Minimax forecast - min the worst-case (max possible) loss

→ Minimax regret - min excess risk over the θ-optimal forecast

Minimax forecast

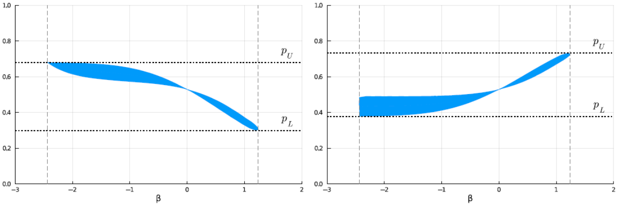
$$d_{q,mm} = \begin{cases} p_U & \text{if } p_U \leq 1/2 \\ p_L & \text{if } p_L \geq 1/2 \\ 1/2 & \text{otherwise} \end{cases}$$

Minimax regret forecast

$$d_{q,mmr} = \frac{(p_L + p_U)}{2}$$

Robust Forecasts

→ Range of forecast probabilities as a function of β



(a) $Y_{iT} = 0$ (b) $Y_{iT} = 1$

Figure 1: Panel probit example with $T = 2$ and $\beta_0 = 0.2$. Shaded regions denote the sets $\{\mathbb{P}_\theta(Y = 1) : (\beta, \Pi_{\lambda,y}) \in \Theta_0\}$ as a function of β. Black dotted lines denote p_L and p_U.

→ $p_L = 0.2997$ and $p_U = 0.6803$ (left panel)

→ Minimax forecast - min the worst-case (max possible) loss is 0.5.

Bayesian/Efficient Robust Forecasts

- → Estimate the set Θ_0 prior to making a forecast.
- → Express Θ_0 as a set-valued function of reduced from parameter P.
- → Specify a prior over P, learn from the data (posterior $\Pi(P|X)$).
- → Bayesian/efficient robust forecasts min integrated maximum risk/regret, which averages over P and the data.
- → Minimax forecast

$$d_{q,mm}(X) = \begin{cases} \int p_U(P) d\Pi & \text{if } \int p_U(P) d\Pi \leq 1/2 \\ \int p_L(P) d\Pi & \text{if } \int p_L(P) d\Pi \geq 1/2 \\ 1/2 & \text{otherwise} \end{cases}$$

- → Evaluate other forecasts by integrated maximum risk/regret, requiring the criterion to be minimized asymptotically. (Efficient plug in — no, but bootstrap of it — yes)

Reactions

→ It is a well-written and precise paper with many detailed and interesting results.

→ I have learned a great deal reading it.

→ The only item on my wish-list is the empirical application. Forecast labor force participation rate to characterize a worse case scenario? Stress-testing is important, could this be a setting applicable to that? Others?

→ Should we be using decision-theoretic framework to deal with model uncertainty in regular macro-forecasting environment?

Reactions

Table 1: Absolute forecast evaluation: uniformity of PIT

	GDP growth		Inflation	
	KS	CvM	KS	CvM
N	**0.93**(**0.51**)	**0.19**(**0.60**)	**1.16**(**0.27**)	**0.40**(**0.23**)
ST^{JF}	**0.94**(**0.52**)	**0.25**(**0.48**)	**0.91**(**0.48**)	**0.30**(**0.32**)
N (ah)	**0.96**(**0.47**)	**0.26**(**0.46**)	1.68(0.06)	0.90(0.05)
ST^{JF}(ah)	**1.03**(**0.40**)	**0.29**(**0.42**)	1.71(0.05)	0.82(0.06)
BVAR	2.29(0.00)	1.86(0.00)	**1.27**(**0.28**)	**0.28**(**0.50**)
PFE	1.72(0.05)	**0.62**(**0.12**)	**1.45**(**0.18**)	**0.53**(**0.18**)
CMM	1.72(0.05)	**0.73**(**0.12**)	**1.44**(**0.17**)	**0.46**(**0.25**)

Note: The table displays the Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) test statistics and p-values of the null hypothesis of uniformity of PITs (in parentheses) for different target variables (in the column headers) and models (in rows). For an explanation of the different abbreviations, see the main text. The p-values are calculated using the block weighted bootstrap proposed by Rossi and Sekhposyan (2019), with block length $\ell = 4$ and 10,000 bootstrap replications. The cases in which uniformity cannot be rejected at the 10% level are reported in bold. The survey dates range from 1997:Q4 to 2017:Q2, with corresponding realizations between 1998:Q3 and 2018:Q1.

Source: Ganics, Rossi and Sekhposyan (2020)