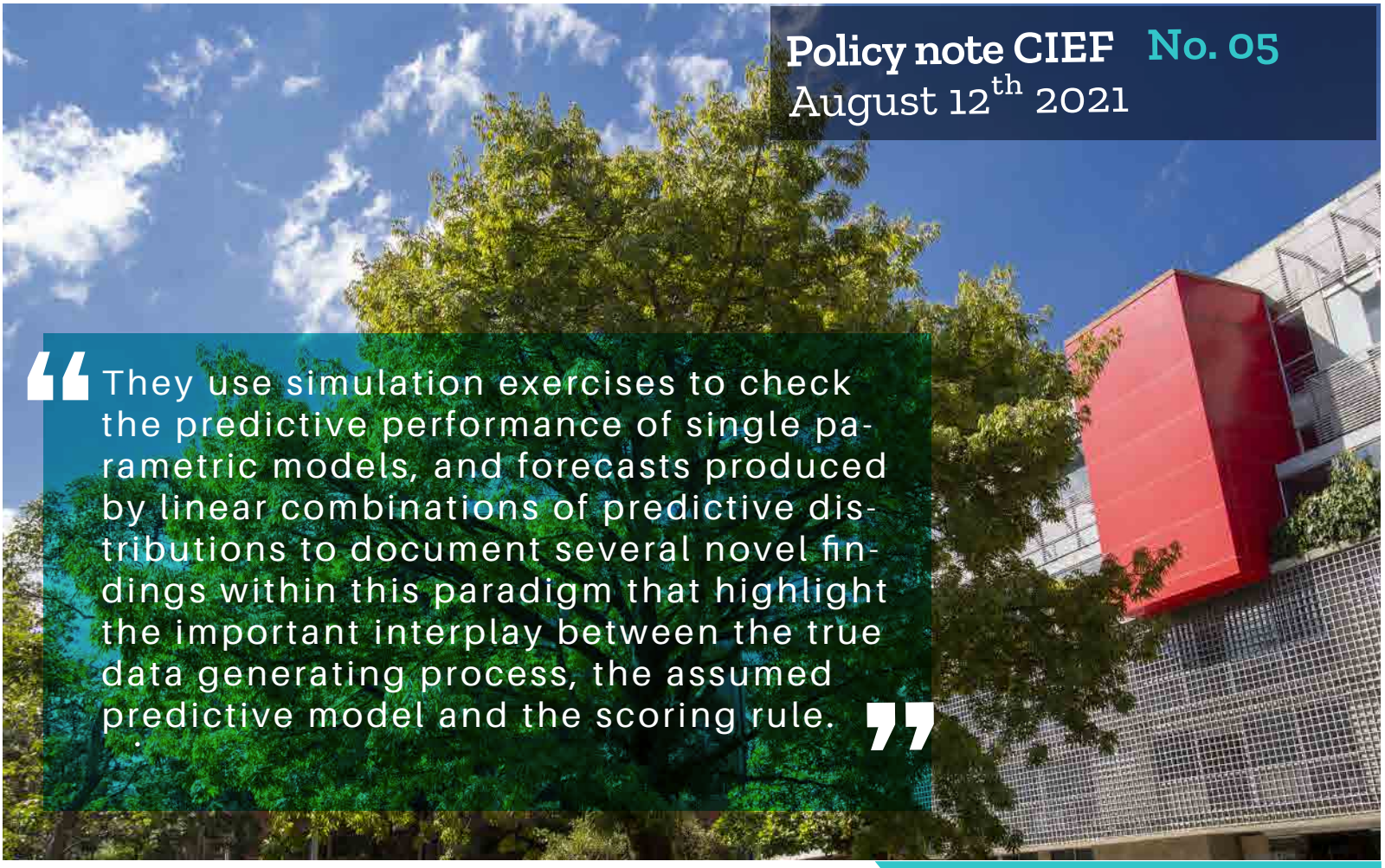
Policy note CIEF No. 05
August 12th 2021



> They use simulation exercises to check the predictive performance of single parametric models, and forecasts produced by linear combinations of predictive distributions to document several novel findings within this paradigm that highlight the important interplay between the true data generating process, the assumed predictive model and the scoring rule.

UNIVERSIDAD EAFIT®
Escuela de
Economía y Finanzas

Optimal probabilistic forecast: when do they work?

Gael M. Martin
Rubén Loaiza-Maya
Worapree Maneesoonthorn
David T.Frazier
Andrés Ramírez-Hassan

Optimal probabilistic forecast: when do they work?

Having probabilistic forecasts, that is, the probability distribution of yet unobserved random variables such as inflation rates, stock prices and gross domestic products can be very useful for policy makers, traders and governments. Proper scoring rules are used to assess the out-of-sample performance of these probabilistic forecasts, with different scoring rules rewarding distinct aspects of forecast performance, for instance, accuracy over specific subsets of the domain of the random variable such as tail behavior or even the whole support. We re-investigate the practice of using proper scoring rules to produce probabilistic forecasts that are "optimal" according to a given score and assess when their out-of-sample accuracy is superior to alternative forecasts, according to that score.

Taking into account that

"Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena."

Box (1976, p. 792)

Particular attention is paid to relative predictive performance under misspecification of the predictive model.

We use simulation exercises to check the predictive performance of single parametric models, and forecasts produced by linear combinations of predictive distributions to document several novel findings within this paradigm that highlight the important interplay between the true data generating process, the assumed predictive model and the scoring rule. Notably, we show that only when a predictive model is sufficiently compatible with the true process to allow a particular score criterion to reward what it is designed to reward, will this approach to forecasting reap benefits. Subject to this compatibility, however, the superiority of the optimal forecast will be greater; the greater is the degree of misspecification. We explore these issues under a range of different scenarios.

We check the validity of our findings in two applications. Financial returns using daily data of S&P500 and MSCIEM index returns where three models are proposed: Naive Gaussian, GARCH and Stochastic Volatility with Jumps, and the optimal linear pool of these three models. First, we observe that strict coherence is in evidence, that is, the optimal forecast under a model and score is strictly preferable to all alternatives, when evaluated according to its own score. Second, it is difficult to assess which model has the best predictive performance overall, due to the interplay between sampling variation and model misspecification. Third, the numerical gains reaped by score-specific optimization in the case of the pool are typically not as large as in the single model cases. Fourth, the equally-weighted predictive pool is outperformed according to all out-of-sample scores by the best performing optimized pool.

We also consider predicting monthly US All Urban Consumers inflation, computed from the Consumer Price Index for All Urban Consumers (CPIAUCSL). We consider three simple models to predict CPIAUCSL: a random walk (RW) model, an ARMA(2,1) model with Gaussian errors, and where the ARMA order was chosen using the Bayesian information criterion (BIC), and a (Gaussian)

linear regression model with inflation, the interest rate, the unemployment rate and the volatility index (VXO), all lagged one period, used as regressors. In addition, we consider an optimal linear pool using these three models. We postulate that all three models are likely to be far too simple to capture all features of monthly inflation, which is typically modelled using multivariate frameworks. We observe that there is no discernible dominance of the score-specific optimal predictive for any out-of-sample measure for the RW model. For the ARMA(2,1) model however, despite the uniform dominance of the LS optimizer, there is more evidence of gains from score-specific optimization. In the case of the linear regression, and for reasons that are not obvious, all out-of-sample measures other than the continuously ranked probability score (CRPS), the preferable optimizer is either censured likelihood score (CLS) 80% or CLS 90%. For the case of the optimized pool, there is a tendency towards strict coherence, with the score-specific optimizer being either the best or second-best performer in most cases.

Our results highlight the care that needs to be taken in the production and interpretation of forecasts designed to be optimal according to a particular measure of forecast accuracy. It is not assured that optimization according to a problem-specific scoring rule will yield benefits; the relative performance of so-called "optimal" forecasts depending on the nature of, and the interplay between the true model, the assumed model and the score. That is, if the predictive model simply does not allow a given score to reward the type of accuracy it should, optimization with respect to that score criterion does not reap benefits.

References

Box GEP (1976) Science and statistics. J Am Stat Assoc 71:791–799

Policy Note CIEF

No. 05, August 12th 2021

Universidad EAFIT

Claudia Restrepo Montoya
Rector

César E. Tamayo Tobón
Dean School of Economics and Finance

Santiago Tobón
Director of the Economic and Financial Research Center
School of Economics and Finance

UNIVERSIDAD EAFIT® Escuela de Economía y Finanzas