Predicting Binary Outcomes Based on the Pair-Copula Construction

Kajal Lahiri and Liu Yang

Department of Economics, University at Albany, SUNY
School of Economics, Nanjing University

ASSA 2021
January 3, 2021

Motivation

- The periodic up-and-down movements in aggregate economic activities, known as business cycles, are of central importance to diverse decision-makings;
- The need for an accurate and timely forecast of future economic conditions arises in the current decisions of manufacturers, investors, financial intermediaries, and central banks alike;
- Since Burns and Mitchell (1946), a great body of literature is devoted to prediction of the underlying evolving state of the real economy.
    - Kauppi and Saikkonen (REStat,2008)
    - Rudebusch and William (JBES,2009)
    - Harding and Pagan (JBES,2011)
    - Berge and Jordà (AEJ,2011)
    - Lahiri and Yang (Handbook,2013)
    - Stock and Watson (JE,2014)
    - Lahiri and Yang (IJF,2015)
    - Lahiri and Yang (JBES,2017)
    - Lahiri and Yang (EL,2016)

Motivation

What is special when predicting binary events?

- Discrete support requires non-linear specification;
- Predictive loss function is of special form.

What characterizes a good binary forecast?

- Traditional measure: mean squared errors (MSE) or Brier's score, i.e. $1/T\sum_{t=1}^{T}(Z_t - P_t)^2$, where P_t is the probability forecast of the event $Z_t = 1$;
- MSE can give a completely misleading conclusion in the case of rare events.

Motivation

To have a high discriminatory power, two conditional distributions must be separated from each other very well.

- The centers of two distributions are far away from each other;
- The variance of each distribution is relatively small.

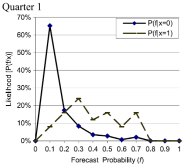
Motivation

Problem with Brier's score (cf. Lahiri and Wang (IJF,2013))

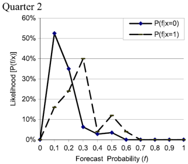
	Q1	Q2	Q3	Q4
MSE	0.0897	0.1065	0.1213	0.1270

Figure: cf. Lahiri and Wang (IJF,2013)

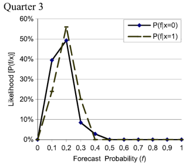
(a) Q1



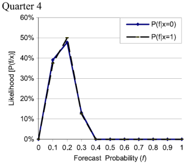
(b) Q2



(c) Q3



(d) Q4



Motivation

Suppose X is a leading index of recession. We predict $Z = 1(\textit{recession})$ iff $X > w$.

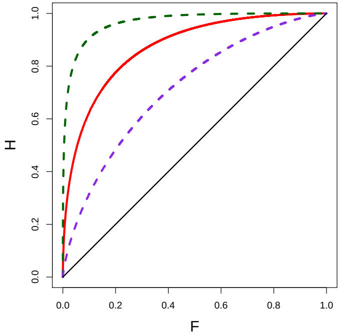
- Hit rate: $H(w) = P(X > w|Z = 1)$;
- False alarm rate: $F(w) = P(X > w|Z = 0)$;
- Both $H(w)$ and $F(w)$ are functions of w.

An ROC curve is generated by plotting the pair $(F(w), H(w))$ in the unit square by varying w.

ROC curve in economic forecasting

- Berge and Jordà (AEJ,2011): NBER recession indicators;
- Lahiri and Wang (IJF,2013): SPF probability forecasts;
- Lahiri and Yang (JBES,2016): Robust confidence bands.

Motivation



Suppose we have I predictors, each of which is denoted by X_i. An interesting question is how to make use of $X \equiv (X_1, X_2, ..., X_I)'$ to maximize hit rate for any given false alarm rate.

Motivation

The most prominent forecasting index in the United States is the Leading Economic Index (LEI) originally constructed by NBER and currently maintained by Conference Board (TCB).

- Simple average of the 10 leading indicators;
- Ad hoc rather than model-based;
- Uncertainty of its optimality in some proper sense.

The purpose of this paper is to improve the accuracy of TCB's LEI in predicting recessions probabilistically.

Optimal rule

If we want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. The two conditional densities of X under H_0 and H_1 are $f(X|\theta_0)$ and $f(X|\theta_1)$ respectively.

Neyman-Pearson lemma

The critical region $C_\eta \equiv \{X \in R^I : \frac{f(X|\theta_1)}{f(X|\theta_0)} > \eta\}$, where η is determined by $P(X \in C_\eta|\theta_0) = \alpha$, is the most powerful test at significance level α.

Suppose $H_0 : Z = 0$ and $H_1 : Z = 1$.

- Power is $P(\frac{f(X|Z=1)}{f(X|Z=0)} > \eta|Z = 1) = H(w)$;
- Size is $P(\frac{f(X|Z=1)}{f(X|Z=0)} > \eta|Z = 0) = F(w)$.

The optimal combination rule to maximize $H(w)$ for any given $F(w)$ (cf. Lahiri and Yang (SNDE,2016)) is

$$\Gamma(X) \equiv \frac{f(X|Z = 1)}{f(X|Z = 0)}. \tag{1}$$

Discussion

Remarks

- Suppose Z is independent of X. (1) is always equal to 1 and the resulting ROC curve is the diagonal. As long as X is of some relevance to predicting Z, the two joint densities in (1) cannot be identical;
- If the conditional distribution of (X_1, X_2) given Z is bivariate normal, (1) is equal to $logit(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2)$;
- If the variances of the two conditional distributions are identical, as is well-known in the discriminant analysis, $\beta_3 = \beta_4 = \beta_5 = 0$. Under this assumption, (1) reduces to a linear encompassing procedure.

Modeling $f(X|Z=1)$ and $f(X|Z=0)$: Vine-copula approach

In our empirical application, each predictor in X reflects one aspect of the whole economy, which is related to future economic activities. TCB uses different transformations to construct these predictors. It is hard to figure out the implied distribution of the transformed predictor.

1st step

$$\hat{f}_i^1(x_i) \equiv \frac{1}{h_i^1 \sum_{t=1}^T I(Z_t=1)} \sum_{t=1}^T I(Z_t=1) k(\frac{x_i - X_{it}}{h_i^1}),$$
$$\hat{f}_i^0(x_i) \equiv \frac{1}{h_i^0 \sum_{t=1}^T I(Z_t=0)} \sum_{t=1}^T I(Z_t=0) k(\frac{x_i - X_{it}}{h_i^0}). \quad (2)$$

The corresponding distribution functions are $\hat{F}_i^1(x_i) \equiv \int_{-\infty}^{x_i} \hat{f}_i^1(u)du$ and $\hat{F}_i^0(x_i) \equiv \int_{-\infty}^{x_i} \hat{f}_i^0(u)du$, both of which are uniform random variables between 0 and 1.

Modeling $f(X|Z=1)$ and $f(X|Z=0)$: Vine-copula approach

A copula is a multivariate distribution function whose univariate marginals are uniforms between zero and one.

Sklar theorem

For any continuous $I-$dimensional random vector X, there exists a unique function C such that $F(x_1, x_2, ..., x_I) = C(F_1(x_1), F_2(x_2), ..., F_I(x_I))$ for any $(x_1, x_2, ..., x_I) \in R^I$. Conversely, $C(v_1, v_2, ..., v_I) = F(F_1^{-1}(v_1), F_2^{-1}(v_2), ..., F_I^{-1}(v_I))$ for any $(v_1, v_2, ..., v_I) \in [0, 1]^I$.

When $I > 2$, only 2 parametric copulas are available:

- Gaussian copula: corresponds to I-dimensional normal distribution;
- t copula: corresponds to I-dimensional t distribution.

Pair-Copula Construction

To fix the idea, we consider $f_{123}(x_1, x_2, x_3)$.

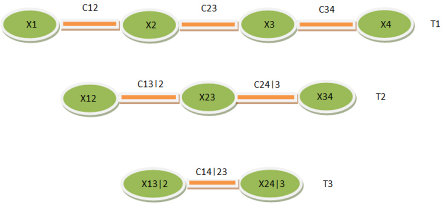
$$
\begin{aligned}
& f_{123}(x_1, x_2, x_3) = f_1(x_1) f_{2|1}(x_2|x_1) f_{3|12}(x_3|x_1, x_2) \\
& = f_1(x_1) f_2(x_2) c_{12}(F_1(x_1), F_2(x_2)) f_{3|12}(x_3|x_1, x_2) \\
& = f_1(x_1) f_2(x_2) c_{12}(F_1(x_1), F_2(x_2)) f_{3|2}(x_3|x_2) c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
& = f_1(x_1) f_2(x_2) f_3(x_3) c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) \\
& c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)). \qquad (3)
\end{aligned}
$$

For $I-$dimensional distribution, PCC consists of I marginal densities, $I-1$ unconditional copulas and $(I-2)(I-1)/2$ conditional copulas (Aas *et al.* (Insurance:M&E,2009)). All copulas involved in PCC are bivariate and thus are relatively easy to parameterize.

Vine copulas: specification

Vine: a nested tree structure to visualize a specific pair-copula construction.

- D-Vine: all predictors are treated equally and none of them plays a central role.



Vine copulas: estimation

Sequential procedure for $I = 3$:

- specify and estimate copula parameters in c_{12} by MLE using data on (X_1, X_2);
- specify and estimate copula parameters in c_{23} by MLE using data on (X_2, X_3);
- compute $F_{1|2}$ and $F_{3|2}$ based on estimates from the previous steps;
- specify and estimate copula parameters in $c_{13|2}$ by MLE using data on $(F_{1|2}, F_{3|2})$;
- take those in previous steps as initial values, and estimate all parameters in c_{12}, c_{23} and $c_{13|2}$ jointly for efficiency purpose.

Optimal Leading Index

For $I = 3$, the optimal combination rule is estimated by

$$\hat{\Gamma}(x) \equiv \frac{\hat{f}_1^1(x_1)\hat{f}_2^1(x_2)\hat{f}_3^1(x_3)\hat{c}_{12}^1(\hat{F}_1^1(x_1), \hat{F}_2^1(x_2))\hat{c}_{23}^1(\hat{F}_2^1(x_2), \hat{F}_3^1(x_3))\hat{c}_{13|2}^1(\hat{F}_{1|2}^1(x_1|x_2), \hat{F}_{3|2}^1(x_3|x_2))}{\hat{f}_1^0(x_1)\hat{f}_2^0(x_2)\hat{f}_3^0(x_3)\hat{c}_{12}^0(\hat{F}_1^0(x_1), \hat{F}_2^0(x_2))\hat{c}_{23}^0(\hat{F}_2^0(x_2), \hat{F}_3^0(x_3))\hat{c}_{13|2}^0(\hat{F}_{1|2}^0(x_1|x_2), \hat{F}_{3|2}^0(x_3|x_2))}$$

Alternatively, a probabilistic leading index of recession can be obtained by

$$L\hat{E}I(x) \equiv \frac{\hat{\Gamma}(x)\bar{Z}}{1 - \bar{Z} + \hat{\Gamma}(x)\bar{Z}} \in (0, 1),$$

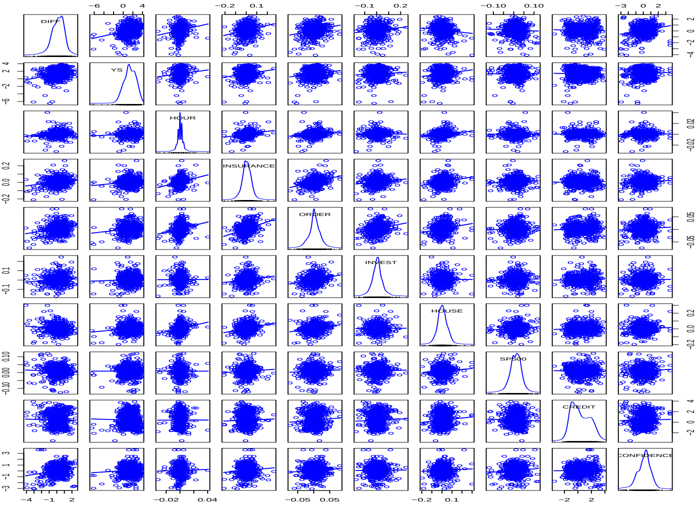
where $\bar{Z} \equiv \sum_{t=1}^{T} Z_t / T$. Since $L\hat{E}I(x)$ is a strictly increasing function of $\hat{\Gamma}(x)$ and thus both yield the same ROC curve.

TCB individual leading indicators

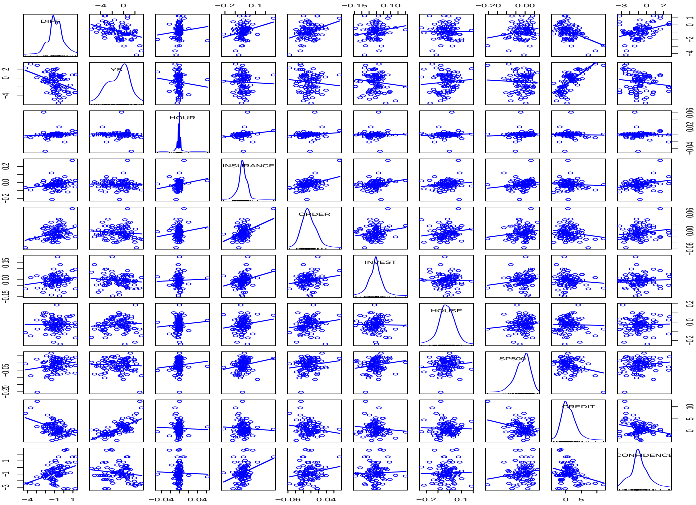
The widely used Leading Economic Index (LEI) proposed by The Conference Board (TCB) is based on the following 10 predictors

- X_1: ISM Index of New Orders;
- X_2: interest rate spread (10-year Treasury bonds less federal funds);
- X_3: average weekly hours (manufacturing);
- X_4: average weekly initial claims for unemployment insurance;
- X_5: manufacturers' new orders;
- X_6: consumer goods and materials, Manufacturers' new orders (non-defense capital goods excluding aircraft orders);
- X_7: building permits, new private housing units;
- X_8: S&P500;
- X_9: Leading Credit Index;
- X_{10}: average consumer expectations for business conditions.

Scatter plot under expansion



Scatter plot under recession



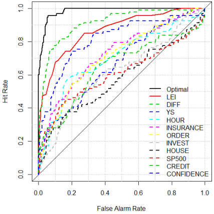
Optimal vine copula

pair	family	parameter	family	parameter
$c_{1,2}$	Gaussian	0.284	270-rotated Clayton	-0.685
$c_{2,3}$	Gumbel	1.087	90-rotated Clayton	-0.247
$c_{3,4}$	Survival Gumbel	1.172	Frank	2.616
$c_{4,5}$	Gaussian	0.386	Gumbel	1.269
$c_{5,6}$	Survival Gumbel	1.198	Gaussian	0.221
$c_{6,7}$	t	0.038(5.828)	Frank	0.624
$c_{7,8}$	Survival Clayton	0.111	Frank	1.036
$c_{8,9}$	90-rotated Clayton	-0.059	90-rotated Joe	-1.291
$c_{9,10}$	Survival Clayton	0.020	270-rotated Joe	-1.355
$c_{1,3\|2}$	Frank	1.134	Gumbel	1.167
$c_{2,4\|3}$	Survival Gumbel	1.083	270-rotated Clayton	-0.083
$c_{3,5\|4}$	Gaussian	0.121	Frank	0.417
$c_{4,6\|5}$	Frank	0.419	Frank	1.876
$c_{5,7\|6}$	Frank	1.581	Frank	1.516
$c_{6,8\|7}$	Survival Clayton	0.037	Survival Clayton	0.321
$c_{7,9\|8}$	Gaussian	0.109	Clayton	0.130
$c_{8,10\|9}$	Gumbel	1.077	Clayton	0.136

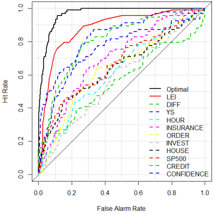
With 200 bootstraps, the calculated p-value for the chosen vine copula is 0.26 (0.4) for expanionary (recessionary) periods (cf. Schepsmeier(ER,2016)).

Figure: In-sample ROC curves for 5 horizons

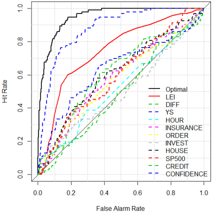
(a) 1 month ahead



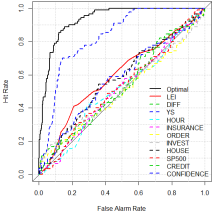
(b) 6 months ahead



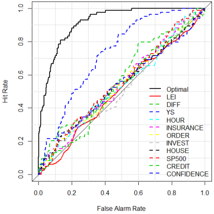
(c) 12 months ahead



(d) 18 months ahead



(e) 24 months ahead



Time series of 6-month-ahead predictive probabilities

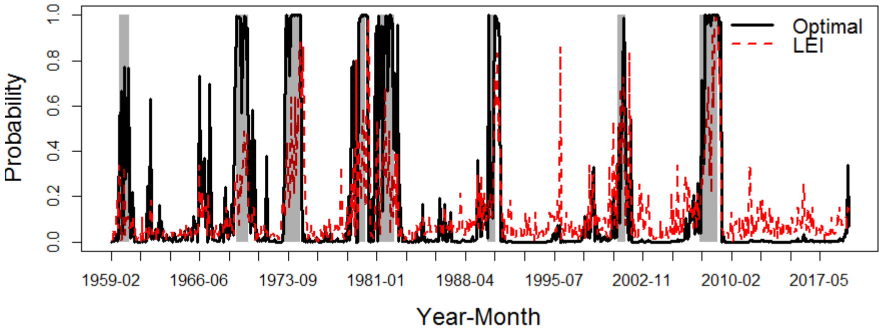
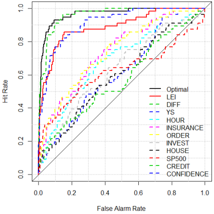
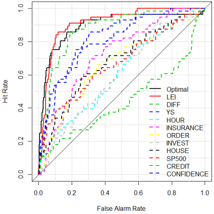


Figure: Out-of-sample ROC curves for 5 horizons

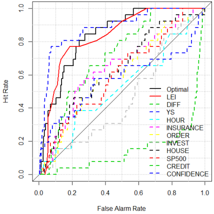
(a) 1 month ahead



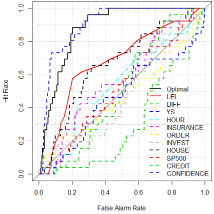
(b) 6 months ahead



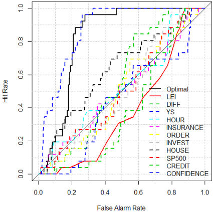
(c) 12 months ahead



(d) 18 months ahead



(e) 24 months ahead



Generalized additive models

- In the preceding analysis, all conditional copulas in PCC are required to depend upon the conditioning variables up to the two conditional marginals—the so called "simplifying assumption";
- Vatter and Chavez-Demoulin (JMA,2015) assumed a conditional dependence measure in a semiparametric manner,

$$\Psi(x;\theta) = g\{z\beta + \sum_{k=1}^{K} h_k(t_k)\}, \tag{4}$$

  where
  - $g : R \to P$ is a strictly increasing and links the expression of generalized additive model with the dependence measure;
  - $z \in R^p$ and $t \in R^K$ are subsets of x or products thereof to consider interactions;
  - $h_k : T_k \to R$ are smooth functions supported on closed $T_k \subset R$ for all k.

Generalized additive models

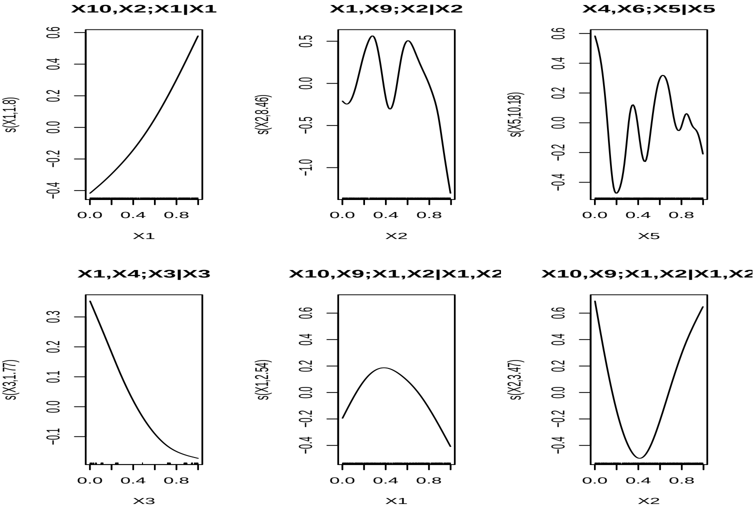
- $\theta \in \Theta$ is the vector of stacked parameters, containing both $\beta \in R^p$ and h_k for all k;
- h_k is smooth and admits a finite-dimensional basis-quadratic penalty representation;
- The penalized likelihood estimators $\hat{\theta}$ maximize

$$l(\theta, \gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} log(c(u_t; \Psi(x_t; \theta))) - \frac{1}{2} \sum_{k=1}^{K} \gamma_k \int_{T_k} h_k''(t_k)^2 dt_k. \quad (5)$$

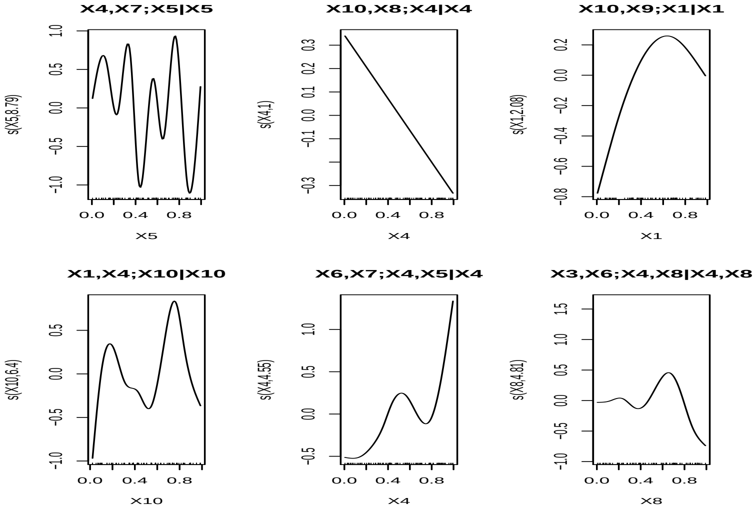
Generalized additive models

- R provides gamCopula package to implement this methodology;
- To save parameters, all conditional bivariate copulas are set to be independent from the fifth level of our vine tree;
- For other conditional copulas, the dependence upon the conditioning variables can be shown in the two conditional plots in both regimes.

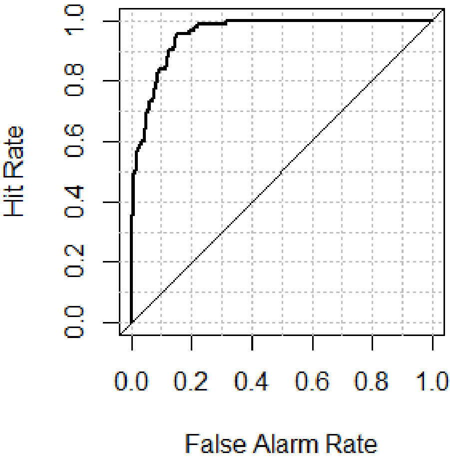
Conditional dependence measure under expansion



Conditional dependence measure under recession



ROC curves based on gamCopula



Simplified- v.s. non-simplified PCC

- Accounting for the variation of Kendall's τ across the range of x_2 is of little value from the perspective of forecasting a binary outcome;
- We do not lose much forecast precision by relying on the simplifying assumption as long as the ratio of Kendall's τ across regimes is close to a constant when the conditioning variable x_2 lies within a segment with high probability mass;
- Considering the computational burden of the conditional PCC model, we prefer using the simplified PCC model in constructing the leading economic index.