GINI METHODOLOGY APPLIED TO MACROECONOMIC FORECASTING

TECHNICAL REPORT

Zhenyu Xu
School of Data Science
Fudan University
18307100053@fudan.edu.cn

Youzhe Liu
School of Data Science
Fudan University
18307100052@fudan.edu.cn

March 10, 2021

ABSTRACT

This paper mainly compares the difference between the Gini method and general linear regression in the performance of prediction, and the predictive ability of the Gini method when applied to complex models for financial data.The dynamic factor model is used as the main support of the model, and the Gini method is used as the modification.Data were collected from U.S. Monthly Macroeconomic Indicators from 1967-12-01 to 2020-10-01. Important models presented in this paper are: autoregressive model, vector autoregressive model, and dynamic factor model

Keywords Gini regression, Factor model, Auto regression

1 Introduction

A large number of methods have been applied to the analysis and forecasting of macroeconomic data. These long-term or short-term forecasts and scenarios screen out the important factors, making it possible to find the inherent logic from a small amount of macroeconomic data. The number of variables with strong correlation with the predicted target data is usually small, so the number of variables used in most forecasting models is often small. The model data categories used in this report tend to be fewer than ten, and fewer remain after simplification.

The widely known Gini coefficient is part of the Gini methodology. The core of the Gini methodology used in this paper is Gini regression. Gini regression is derived from the use of Gini Mean Difference(GMD) which is seen as a measure of variability. The Gini mean difference describes the nature of the distribution of variables, and can also describe the relationship between variables, namely the Gini correlation coefficient, which is similar to correlation coefficient. The regression and test methods based on this are also of great value in application. Compared to the general linear regression under the hypothesis of normal distribution, Gini Regression allows variables to have larger randomness, which corresponds to the strong irregularity of economic data. But few scholars have used the Gini method in actual macroeconomic forecasting, and fewer have combined the Gini method with complex models to test the forecasting effect. This paper will study the performance of the model by combining the Gini method with other benchmark models. In practice, commonly used models for predicting time series include: autoregressive model (AR) and its extended vector autoregressive model (VAR), moving average model (MA) and other time series models, as well as dynamic factor model (DFM) and approximate dynamic factor model (ADFM). Most of these models have performed well in practical applications, but are only based on standard linear space projections. Gini method provide another projection and regression method and will be tested in this report. We will study the comparison with Gini method and ordinary linear regression and focus on the performance of AR, VAR, DFM which are all combined with Gini regression or Gini correlation coefficient.

2 General Multiple Regression Task

2.1 Data source and data processing

Measuring an economy and predicting its future trajectory relies on analyzing key pieces of macroeconomic data. Here, we take a look at some of the most important indicators that measure everything from economic growth to price changes to unemployment. Data is download in website fred.stlouisfed.org, FRED economic research center from 1967-12-01 to 2020-10-01 monthly. The data can reflect macroeconomic developments and can be a indicator for the whole country's economic level, which is a very important tool to analysis the economics of an area.

Table 1: Variable List

Variable name	Complanation
INDPRO	Industrial Production: Total Index:
W875RX1	Real personal income excluding current transfer receipts
CMRMTSPL	Real Manufacturing and Trade Industries Sales
AYEMS	All Employees, Total Nonfarm
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items in U.S. City Average
PCEPI	Personal Consumption Expenditures: Chain-type Price Index
COREFLEXCPIM159SFRBATL	Flexible Price Consumer Price Index less Food and Energy
WPUFD49207	Producer Price Index by Commodity: Final Demand: Finished Goods

The overall economic situation showed an upward trend. In the following report, we focused on the model fitting or prediction rather than the actual significance of the model obtained by fitting.

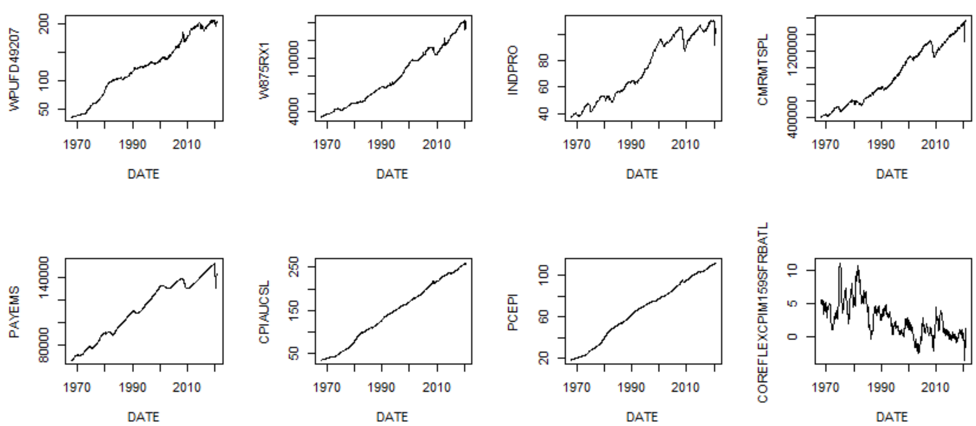

Figure 1: All the trendings

We forecast changes in the first four economic indicators. As use an example of INDPRO.

Current data need some transformation (Including but not limited to):

1. Diffusion Index transformation

The diffusion index is used to measure the rate of change and detect economic turning points. Forecasting continuous data is often not accurate, and we often prefer to know whether economic data is increasing or decreasing.

$$X_t^D = \frac{X_t - X_{t-1}}{X_t}$$

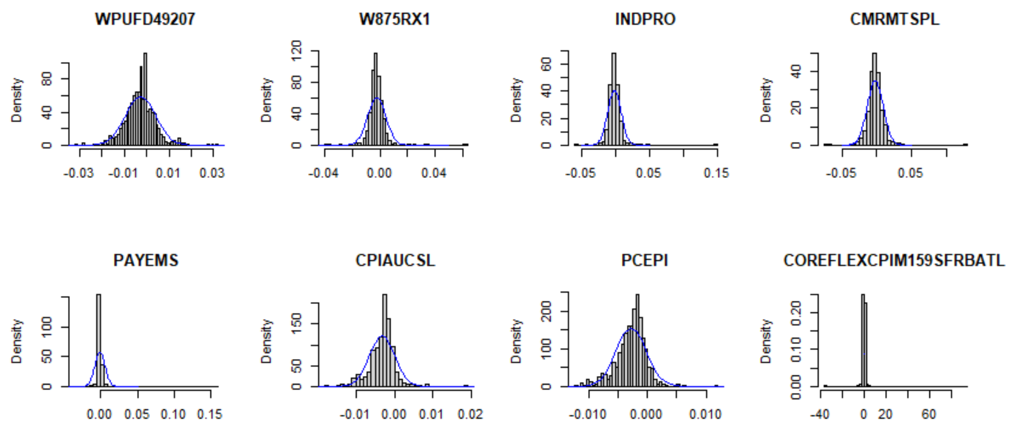

Figure 2: All the Data Changing Rate

2. Modified Principle Components Analysis(PCA)

General principal component analysis will choose to decompose the covariance matrix, but in the later model, we choose to decompose the Gini covariance matrix.

$$F = XQY = FA$$

3. Lag extension (time series transformation)

The number of lagged items can be selected by the following criteria: AIC (Akaike Information Criterion), HQ (Hannan-Quinn Criterion), SC (Schwarz Criterion) FPE (Final Prediction Error Crite-RIA). We will refer to the four criteria to select the appropriate number of lagging items.

$$X_t \to X_{t-1}, X_{t-2}, ..., X_{t-p}$$

4. Order statistics (applied with Gini)

In the process of using Gini regression, order statistics play an obvious role in the process of calculating covariance matrix.

$$X \to F(X)$$

2.2 Comparison between linear regression and Gini regression

For the general regression method, the predictive variables and independent variables meet the following assumption:

$$\operatorname{cov}(Y, T_i) \equiv \sum_{j=1}^{K} \beta_j \operatorname{cov}(X_j, T_i) + \operatorname{cov}(\epsilon, T_i), i = 1, 2, .., K$$

When $T_i = X_i$, the coefficients solved are the same as those solved by the least square linear regression. When $T_i = F(X_i)$, it represents the semi-parametric Gini regression (ordinary Gini regression) and $T_i = -[1 - F(X_i)]^{v_k}$ for extended Gini regression. This means that on the projection space OLS projects the original data onto the original coordinate axis, whereas the Gini regression projects the order statistics of variables. So the Gini regression residuals are orthogonal to the order statistics rather than to the original independent variables. Extended Gini regression is more complicated, and it's equal to semi-parametric Gini regression when $v_k = 1$.

We use an example of fitting to illustrate the difference between these methods. The diffusion index of the data fluctuates greatly and it's difficult to see whether the fit is good from the image. Consequently after fitting we sorted the real data by order. The graph below describe the residuals after being fitted by two method. In all, the difference is not so obvious.

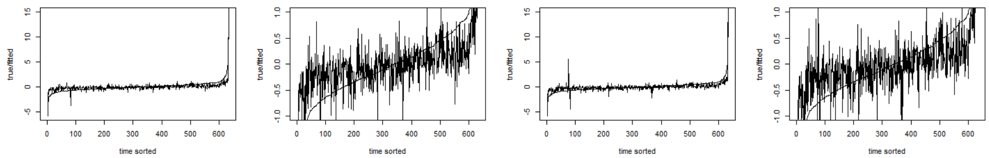

Figure 3: Comparison of fittings(Left: OLS Right: Gini)

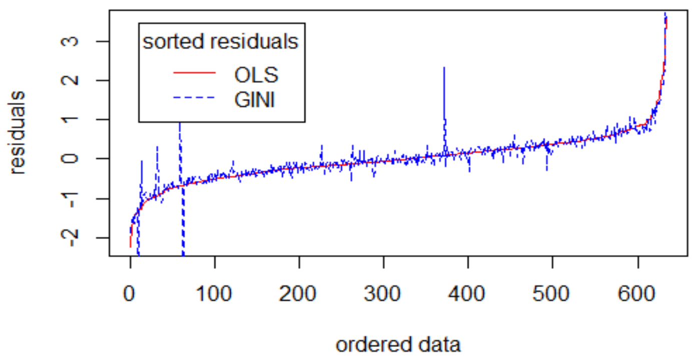

Figure 4: Order of Error

There is not much difference in the solutions, but it can be seen that the Gini regression produces more outliers. The sensitivity of the two methods to outliers can be seen by adding the random error term to judge the change distribution of the coefficient. As the number of error samples increases, the Gini regression will be the first to reach the limit of change, but OLS will continue to be affected (though to a lesser extent).

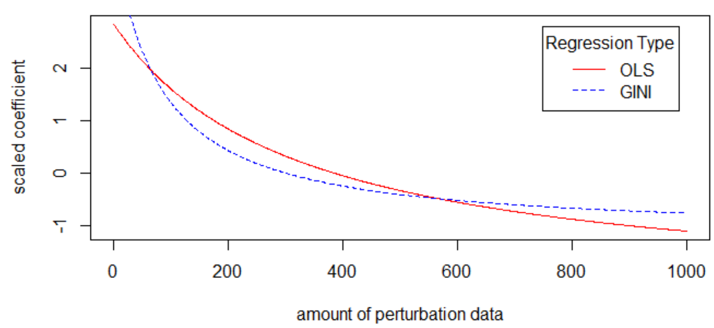

Figure 5: Distribution of disturb

But another significant difference between the two methods is that the pre-existing influence between the Gini regression data is greater than that of OLS. The existence of order statistics makes the initial data may be greatly affected when a large number of samples flood in (but this is generally not the case in financial data). When the order statistics are affected either a small amount of interfering data or a large amount of interfering data can have a huge impact on the results .

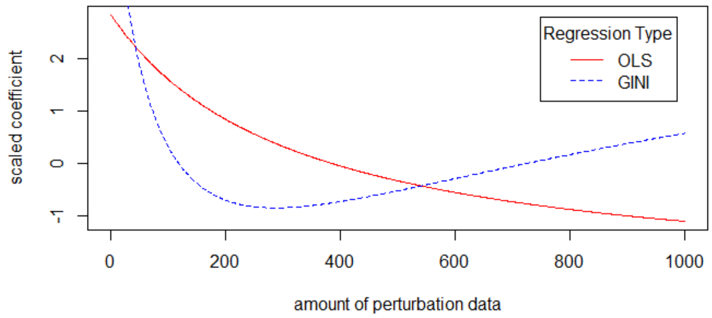

Figure 6: A Special Case

There are significant differences in the performance of these two methods. Under the general normal assumption of residual error, OLS is obviously better than Gini regression. In fact, OLS is probably the best regression under normal assumptions. However, although the Gini distribution does not perform well under the normal distribution hypothesis, the Gini distribution does perform well under some heavy tail distribution conditions, such as generating residual distribution, perhaps due to the insensitivity of Gini regression to large residuals.

$$f\left(y\right)=k\left(\nu\right)\exp\left\{-\frac{1}{2}|\frac{y}{\lambda_{v}}|^{\nu}\right\}\lambda_{\nu}=\left\{\frac{2^{-\frac{2}{\nu}}\Gamma\left(v^{-1}\right)}{\Gamma\left(\frac{3}{\nu}\right)}\right\}^{-\frac{1}{2}},k\left(\nu\right)=\frac{\nu}{\lambda_{\nu}2^{1+\frac{1}{\nu}}\Gamma\left(v^{-1}\right)}$$

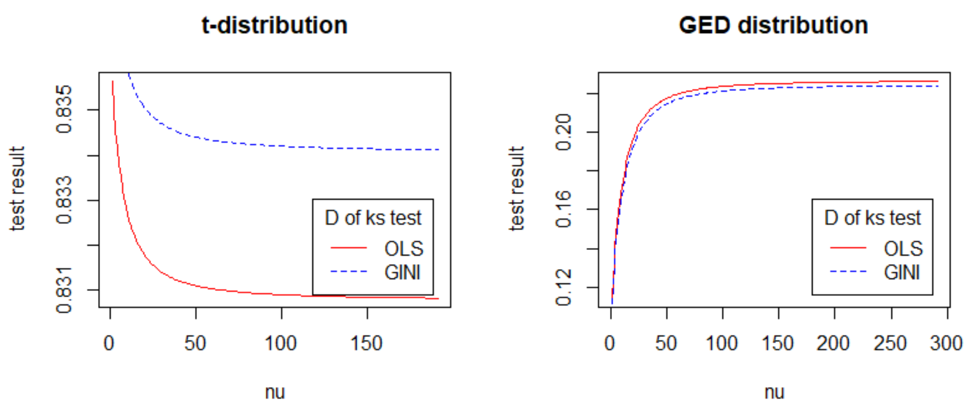

Figure 7: KS testing

3 Gini Regression Applied to advanced economics models

Gini regression is essentially the choice of projection mode, so we can combine the Gini method with the higher-order model by changing the projection mode in the model. The results of these models using the Gini method and the linear regression method have little difference, and the performance is mainly different under different model assumptions. In order to distinguish these models, we take the mean square error and the likelihood of generating residual distribution as the test and evaluation criteria of the models

3.1 Auto Regression And Vector Auto Regression

Autoregression uses observations from previous time steps as input to a regression equation to predict the value at the next time step. Basically similar to the commonly used autoregressive model, the length of lag items is selected through ACF or PACF and then the Gini regression is used for fitting. However, we also use the lag term of independent variables in the prediction results to carry out more accurate fitting under the premise of controlling the lag order between 1 and 6.

Subsequent results show that the vector auto regression performed well under the GINI regression.

Description of ordinary auto regression:

$$Y_t = \sum_{i=0}^{p_1-1} A_i X_{t-i} + \sum_{j=1}^{p_2} k_j Y_{t-j} + \epsilon_t$$

Vector can capture the relationship between multiple quantities as they change over time, and we will forecast these quantities at the same time to show their relationship. But the model turned out to be underperforming.

For more complicated vector auto regressions:

$$Y \in R^{N\times T}$$
$$Y_t = \sum_{i=0}^{p_1-1} A_i X_{t-i} + \sum_{j=1}^{p_2} B_j Y_{t-j} + \epsilon_t$$

3.2 Economics Framework: Approximate Dynamic Factor Model

The factor model can be described as the following equations:

$$Y_t = \sum_{i=0}^{p_1-1} A_i F_{t-i} + \sum_{j=1}^{p_2} k_j Y_{t-j} + \epsilon_t F_t = Q X_t$$

The forecasting procedure are of two steps: (1) The sample data X_t are used to estimate a time series of factors, $\hat{F}_t$ (2) We can obtain the coefficients by regression y_{t+1} onto variables.

The common factor calculation method is to use the covariance matrix for singular value decomposition, but the projection method used by the Gini regression is different from the projection method of simple linear regression, so we choose the Gini covariance matrix for singular value decomposition in practical operation.

$$\mathrm{Gcov}\,(x, y, \nu) = -\nu \times \left(x, (1 - F\,(y))^{\nu-1}\right)$$
$$\Gamma_{X,Y,\nu=1} = \frac{\mathrm{cov}(X, G(Y))}{\mathrm{cov}(X, F(X))}$$
$$A_G = U \Sigma V$$

3.3 Forecasting Models and Data

The data is the same as the one in regression comparison above, which includes 8 variables for macroeconomics.

Using the Gini regression as the benchmark model and the above three models combined with the Gini regression, we split the US macro data into the training set and the test set to make predictions and calculate the corresponding mean square error and likelihood based on the generated residual distribution.

col: Gini multiple regression // Univariate auto regression // Vector auto regression // Approximate Dynamic Factor Model

row: Forecast on

- INDPRO: Industrial Production: Total Index:

- W875RX1: Real personal income excluding current transfer receipts

- CMRMTSPL: Real Manufacturing and Trade Industries Sales

- PAYEMS: All Employees, Total Nonfarm.

The monthly and yearly data are shown below:

Table 2: Monthly MSE and GED

	MSE				GED			
	GMR	UAR	VAR	ADFM	GMR	UAR	VAR	ADFM
INDPRO	5.866161	9.816599	6.780173	6.051836	2132.478	1706.521	1659.699	2094.005
W875RX1	5.879104	9.893108	5.283107	5.829427	2134.188	1748.056	1619.183	1993.565
CMRMTSPL	5.964107	11.859183	11.542550	10.923472	2151.817	1829.892	1838.224	2261.859
PAYEMS	5.838686	10.692358	6.453117	5.774791	2121.398	1733.873	1633.345	1949.537

Table 3: Yearly MSE and GED

	MSE				GED			
	GMR	UAR	VAR	ADFM	GMR	UAR	VAR	ADFM
INDPRO	3.721765	8.395832	1.867337	3.172965	116.6358	187.8207	64.32403	104.24799
W875RX1	3.536865	9.357921	1.669219	2.741249	120.0233	190.4945	57.48455	90.20577
CMRMTSPL	2.954295	10.211061	5.700357	8.024650	104.6982	205.8162	139.37212	186.48225
PAYEMS	2.810709	7.025152	2.783390	2.995471	[110.2309	175.6094	93.74296	98.09155

On the whole, it is clear that the performance of dynamic factor model is better than other models, but there are differences in different methods when predicting some variables.

In addition, the results obtained by the decomposition of the Gini correlation matrix in the principal component analysis are obviously better than the decomposition of the covariance matrix. This is because the approximate dynamic model adopts the Gini regression rather than simple linear regression, so the Gini matrix contains more explanatory information for variables in the projection space.

4 Conclusion

We find three features of the empirical results intriguing. There are usually no more than four factors that play an important role in the model, and the effects among the factors are interactive. The factors that play a prominent role in the model are often easier to be predicted, especially the effect predicted by the approximate dynamic model is better.

Moreover, the number of lagged items in the time series is usually no more than two, and the change of the series is usually positive correlation, which indicates that the frequency of continuous increase or decrease is very high but the time is not long, no more than three consecutive months (a quarter), the number of lagged items in the annual data is less, and the degree of autocorrelation of the series is lower.

In conclusion, Gini regression method is a powerful tool to solve some specific tasks which is not a standard normal distribution. Gini method can be applied to some occasions based on some advanced economics models can achieve quite good effects.

Reference

Shlomo Yitzhaki and Edna Schechtman: The Gini Methodology. Springer, New York, NY.

Stock, James H and Watson, Mark W, 2002. "Macroeconomic Forecasting Using Diffusion Indexes," Journal of Business & Economic Statistics, American Statistical Association, vol. 20(2), pages 147-162, April.

Hiro Y. Toda and Peter C. B. Phillipsă(1994)ăVector autoregression and causality: a theoretical overview and simulation study,ăEconometric Reviews,ă13:2,ă259-285