WILEY | Hindawi

Research Article

A Comparison of Autometrics and Penalization Techniques under Various Error Distributions: Evidence from Monte Carlo Simulation

Faridoon Khan,[1] Amena Urooj,[1] Kalim Ullah,[2] Badr Alnssyan,[3] and Zahra Almaspoor [4]

[1]*PIDE School of Economics, Pakistan Institute of Development Economics, Islamabad, Pakistan*
[2]*Foundation University Medical College, Foundation University, DHA-1, Islamabad, Pakistan*
[3]*Department of Administrative Sciences and Humanities, College of Community at Buraydah, Qassim University, Buraydah 51452, Saudi Arabia*
[4]*Department of Statistics, Yazd University, Yazd 89175-741, Iran*

Correspondence should be addressed to Zahra Almaspoor; z.almaspoor@stu.yazd.ac.ir

This work compares Autometrics with dual penalization techniques such as minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) under asymmetric error distributions such as exponential, gamma, and Frechet with varying sample sizes as well as predictors. Comprehensive simulations, based on a wide variety of scenarios, reveal that the methods considered show improved performance for increased sample size. In the case of low multicollinearity, these methods show good performance in terms of potency, but in gauge, shrinkage methods collapse, and higher gauge leads to overspecification of the models. High levels of multicollinearity adversely affect the performance of Autometrics. In contrast, shrinkage methods are robust in presence of high multicollinearity in terms of potency, but they tend to select a massive set of irrelevant variables. Moreover, we find that expanding the data mitigates the adverse impact of high multicollinearity on Autometrics rapidly and gradually corrects the gauge of shrinkage methods. For empirical application, we take the gold prices data spanning from 1981 to 2020. While comparing the forecasting performance of all selected methods, we divide the data into two parts: data over 1981–2010 are taken as training data, and those over 2011–2020 are used as testing data. All methods are trained for the training data and then are assessed for performance through the testing data. Based on a root-mean-square error and mean absolute error, Autometrics remain the best in capturing the gold prices trend and producing better forecasts than MCP and SCAD.

1. Introduction

In the regression analysis, it is the core concern of researchers to discover the key predictors for achieving better prediction of the response variable. Therefore, identifying the potential predictors for knowledge discovery and boosting the predictive power of the model are very beneficial [1]. However, to construct a linear regression model, variable selection is one of the most vital steps. In practice, a large number of predictors can raise the variance of the fitted model, and selecting several predictors may result in unpredictable output or biased results. In other words, incorporating more predictors in the model may cause high

variation in the least-squares fit, which, in turn, results in overfitting the model, and hence, it yields a poor forecast for the future [2]. Furthermore, if the predictors are highly correlated with each other, then the standard error associated with each regression coefficient tends to increase, which leads to invalid inferences [3–5]. On the other hand, missing a single important predictor may lead to model mis-specification, and the conclusion drawn on the basis of a particular model could be misleading [6].

In the recent era, a considerable chunk of research is focused on the analysis of "high-dimensional" data in the discipline of finance and economics. Resultantly, a considerable focus is being paid to the varieties of techniques

that are applicable in the domain of data mining, dimension reduction, and machine learning [7, 8]. Among them, penalization techniques and Autometrics are very popular to handle huge data sets [9].

Many studies exist in the literature in which the performance of Autometrics is determined theoretically as well as empirically. Some of them are [9–16]. Similarly, many researchers have evaluated the penalization techniques under time series set up such as Mol et al. [17], Inoue and Kilian [18], Bai and Ng [19], Kim and Swanson [20, 21], Luciani [22], Swanson and Xiong [8, 23], Swanson et al. [24]; and Maehashi and Shintani [25].

In the above papers, penalization techniques are often compared to each other, and just a few papers have compared the Autometrics with penalization techniques such as least absolute shrinkage selection operator (Lasso), adaptive Lasso, and weighted adaptive Lasso. To date, none of the papers has considered the modified form of penalization techniques in our context. Hence, this study contributes in two dimensions. Firstly, we consider two modified penalization techniques: minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) and compare with Autometrics theoretically as well as empirically. Secondly, the comparison is made under asymmetric error distributions instead of Gaussian.

Our study aims to compare Autometrics with improved penalization techniques including smoothly clipped absolute deviation and minimax concave penalty under several asymmetric error distributions such as exponential, gamma, and Frechet through Monte Carlo simulations. Moreover, we alter the sample size, number of predictors, and magnitudes of multicollinearity in order to determine their effect on the considered techniques. For real phenomenon analysis, we consider a financial data set.

The remaining part of the work is organized in the following way. In Section 2, we have elaborately discussed the model selection techniques and data-generating process. Monte Carlo evidence on the comparative performance of various model selection methods is discussed in Section 3. Real data applications are described in Section 4. Section 5 gives the concluding remarks.

2. Model Selection Techniques

Model selection is one of the crucial steps of empirical research throughout all disciplines, where an earlier theory does not predefine a complete and correct specification. Economics is certainly one of them, as macroeconomic processes are typically high-dimensional, nonstationary, and complicated [26]. Commonly, many different solutions have been recommended to fit the data. Hence, statistical model selection becomes a primary and ubiquitous task in empirical economic research.

Selection procedures such as information criteria, stepwise, and penalized regression are unavoidable. There can never be a consensus regarding which model is best because there is a considerable amount of criteria to assess the model's performance. Luckily, during the past two decades, a new revolution has been existing in model building,

in the form of general-to-specific modeling, indicated by gets, as contained in the computer program, named as PcGive. Computer automation of gets methods has shed light in a new way on the statistical model selection.

PcGive is a computer program that automatically selects an econometric model. It is absolutely a new approach to formulate models and particularly devised for handling economic data when the correct form of an equation under analysis is unknown. In PcGive, the automatic model selection job is performed by Autometrics. Hence, in the next section, we provide a detailed explanation of Autometrics.

2.1. Autometrics. The automated gets procedure is almost be considered a "black box": a final model is chosen from the model that is constructed from an initial set of candidate variables. The initial model refers to the general unrestricted model (GUM). Mostly, a set of terminal candidate models is found. In such circumstances, information criteria are utilized as the tiebreaker. There is a possibility that we may choose the final GUM in the block-search procedure, which is the union of the terminal candidate models.

The aim of the automated gets procedure is that the GUM is well specified statistically, which is subjected to misspecification testing. Hereafter, diagnostic tests guarantee that all underlying terminal candidate models clarify these tests as well. Simplication of GUM is done via path search. Such a type of search is needed to tackle the complex autocorrelation that is often present in macroeconomic data. A simplification is acceptable provided the expelled variables are insignificant and the new model is a sound chopping of the GUM. The latter condition is also known as encompassing the GUM or backtesting and, in the context of linear regression models, is based on the F-test of the removed variables.

In the application of Autometrics, reduction in p-value p_α is the principal choice to be used for backtesting and individual coefficient significance. There are some tools to eschew estimating 2^k models [27]. This method is very efficient even though the costs of statistical inference cannot be circumvented and the costs of searching are substantially low. A pair of automatic model selection frameworks that fail to fit the model within general-to-specific (gets) methodology are as follows:

(1) Stepwise regression: start with the empty model and add the most significant omitted variable in the model. The highly insignificant variable is removed from the model that is observed at any stage . Hence, in every iteration, we include one significant variable and discard an insignificant variable [28]. This method is repeated till we get all the variables in the model to be significant, and all omitted variables must be insignificant.

(2) Backward elimination: all predictors are entered into the initial model; then predictors are thrown one at a time starting from the least significant. The process is continued until all predictors have a p-value of p_a or small.

There exist three main differences with automated gets: (i) lack of search, (ii) no backtesting, and (iii) no mis-specification

testing/diagnostic tracking. Figure 1 describes the way that how Autometrics selects the model automatically.

2.1.1. Methodology.

Autometrics comprises the following five basic stages:

(i) In the first stage, the linear model known as the so-called general unrestricted model (GUM) is formed

(ii) In the second stage, the parameters are estimated along with testing the statistical significance of the GUM

(iii) In the third stage, the presearch process is performed

(iv) The fourth stage produces the tree-path search

(v) In the last stage, the final model is selected

Doornik [27] elaborated the entire algorithm of Autometrics whereas the steps to run Autometrics are as follows. Start off by considering all the candidate variables in a linear model (GUM), estimate it by the least-squares method, and then verify through diagnostic tests. In case of insignificant coefficients, then simpler models are estimated utilizing a tree-path reduction search and validated by diagnostic tests. If some terminal models are detected, Autometrics undertakes their union testing. Rejected models are deleted, and the union of those terminal models who survived induces new GUM for another tree-path search iteration. This inspection procedure continues, and the terminal models are statistically assessed against their union. If two or more terminal models clear the encompassing tests, and then the prechosen information criterion is a gateway to a final decision.

2.2. Shrinkage Methods.

One of the assumptions of the classical linear regression model is that there is no association among the covariates, which often does not exist in practice. If this assumption is violated, then such a phenomenon is known as the problem of multicollinearity. In presence of multicollinearity, it is a challenging task to estimate the reliable effects of a specific covariate. More specifically, the estimated coefficients have high sampling variance along with false signs, due to which both estimation and prediction are poorly affected.

An alternative most used family of methods to deal with many features is the regularization/penalization regression, which includes many methods, but our study selects the most well-known and robust methods: minimax concave penalty and smoothly clipped absolute deviation. A form of the regularized least-squares estimator is the minimizer of the given objective function:

$$\frac{1}{2m} Y - X\gamma^2 + \sum_{k=1}^{p} \pi_k(|\gamma_k|), \tag{1}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)$, $X = (x_1, x_2, \ldots, x_n)$, and γ is the coefficient matrix with $\gamma = (\gamma_1, \gamma_2, \ldots \gamma_p)$. Here, p and m denote the number of covariates and observations, respectively. The second term in equation (1) $\pi_k(|\cdot|)$ represents the penalty function, which adopts different shapes for different procedures.

The term π refers to the tuning parameter that controls the amount of shrinkage. The range of tuning parameters lies between zero and infinity.

We provide the brief discussion of the following methods:

Least Absolute Shrinkage and Selection Operator: the L_1 norm is defined as $\pi_k(|\vartheta|) = \pi|\vartheta|$ tends to the Lasso estimator, where π refers to the tuning parameter and is selected through cross-validation [29]. L_1 norm shrinks the several regressor coefficients to zero retaining the relevant predictors only. norm shrinks the several regressor coefficients to zero retaining the relevant predictors only. If there is a high correlation among the group of predictors, then Lasso keeps only one predictor from the group. In addition, Lasso is biased in features selection [30].

Smoothly Clipped Absolute Deviation: the continuous differentiable penalty function can be defined as: $\pi'_s(|\vartheta|) = s\{I(\vartheta \leq s) + ((\omega s - \vartheta) + /(\omega - 1)s) I(\vartheta \leq s)\}$ If the results of $\omega > 2$, $\vartheta > 0$, and $s > 0$ then the resulting penalty refers to SCAD [31], where $\pi_k(|\vartheta|) = 0$ and $\omega \approx 3.7$ as recommended by Lu et al. [32].

Minimax Concave Penalty: The minimax concave penalty is illustrated as follows: $\pi'_s(|\vartheta|) = ((\omega s - \vartheta) + /\omega)$, where the value of ω is 3.7. This procedure provides the convexity of the penalized loss in sparse regions substantially given certain thresholds for variable selection and unbiasedness [33].

2.3. Selection of Tuning Parameter.

The tuning parameter λ is often selected using a cross-validation approach aimed at achieving the optimum prediction solution. It entails splitting the given data into two halves at random: a training data set and a testing data set (or holdout set). The training data set is being used to fit the model, and the fitted model will be used to anticipate the responses for the validation set data. The test error rate is estimated by the validation set error rate, which is commonly calculated using MSE in the context of a numerical response. The k-fold cross-validation method involves randomly splitting data collection into k groups, or folds, of roughly similar size, using a k-fold CV; usually, we use k that is equal to 10 or 5. The algorithm is fitted on the remaining $k - 1$ folds, with the initial fold serving as a validation set. On the observations in the holdout fold, the mean squared error, MSE_1, is calculated. This technique is repeated k times, with each validation set consisting of a distinct set of observations. MSE_1, MSE_2, $\ldots$, MSE_k are the test error estimates produced by this method. Averaging these values yields the k-fold CV estimate.

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^{k} MSE_i. \tag{2}$$

2.4. Artificial Data-Generating Process.

In the recent section, we introduced some scenarios intending to demonstrate the performance of Autometrics against shrinkage methods
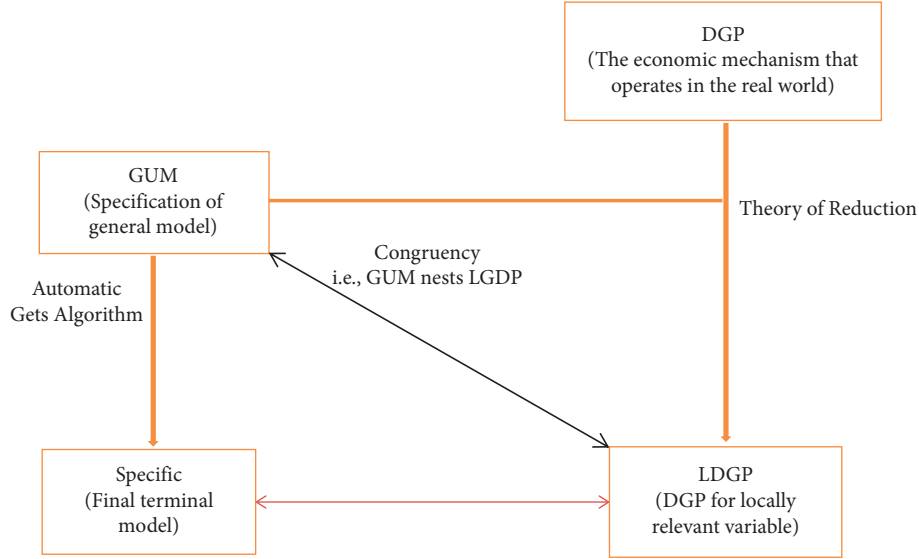
FIGURE 1: From DGP to models.

delineated in previous subsections. We consider two types of correlation structure among covariates, that is, low (0.25) and high (0.90) with varying the distribution of error terms. Our study uses the data-generating process followed by Doornik and Hendry [13] and Wahid et al. [34] to generate artificial data as follows:

$$Y_i = x_i^T \gamma + e_i, \qquad (3)$$

where Y_i is the response variable. The set of covariates, $x_i = x_1, x_2, \ldots, x_P$, is generated from multivariate normal distribution as $x_i \sim MVN(0, \sum)$ where the mean of covariates is zero and $\sum$ is the variance-covariance matrix. It is fact that the variance-covariance matrix contains variance and covariance together. In our case, the variance is assumed to be one, and the covariance between x_m and x_n is generated in the following way: $\text{cov}(x_m, x_n) = \sum^{|m-n|}$ [34].

$$\sum_M = \begin{bmatrix} 1 & . & . & . & \rho^{|n-m|} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho^{|m-n|} & . & . & . & 1 \end{bmatrix}, \qquad (4)$$

which permits for regulation of the degree of pairwise correlation between the covariates m and n by altering the single parameter ρ. Furthermore, γ represent the regression coefficients, and e_i is the disturbance term, which is generated from the following three asymmetric probability distributions in this study. The distributions are exponential distribution, gamma distribution, and Frechet distribution.

The reason behind the selection of these distributions from a huge set of distributions is: exponential distribution is basically a standard distribution in the literature of asymmetric distributions. Moreover, Frechet distribution, which is also known as inverse Weibull distribution, and gamma distribution are the generalized form of the exponential

distribution. Mostly, the distribution of financial data is right-skewed [35].

For our study, we are considering three asymmetric probability distributions from a huge list of distributions. There are many:

(i) $e_i \sim \exp(0, 8)$
(ii) $e_i \sim \text{gamma}(2, 5)$
(iii) $e_i \sim \text{frechet}(1, 2, 4)$

2.4.1. Scenario 1. We perform simulation experiments where we consider three cases of covariates: $p \in \{20, 25, 30\}$. In each experiment, we assume 15 predictors are relevant, and the remaining are irrelevant.

(i) $\gamma = \{(1, \ldots, 1)_{15}, (0, \ldots, 0)_5\}$
(ii) $\gamma = \{(1, \ldots, 1)_{15}, (0, \ldots, 0)_{10}\}$
(iii) $\gamma = \{(1, \ldots, 1)_{15}, (0, \ldots, 0)_{15}\}$

We consider two cases of sample size $n \in \{100, 300\}$. In this scenario, we generate errors of the model from an exponential distribution.

2.4.2. Scenario 2. Furthermore, this scenario is the same as the first experiment; only the errors are generated from a gamma distribution.

2.4.3. Scenario 3. This scenario is the same as the first experiment; besides, the error is generated from the Frechet distribution.

2.5. Measures of Methods Performance. There are a few ways to evaluate the models' performance in terms of variable selection, in which we are adopting the potency and gauge. Gauge is delineated as the empirical null retention frequency

that how often irrelevant covariates are retained. The comparison of Autometrics with penalization methods is evaluated in the form of correct zero identification interpreted as potency and incorrect zero identification referred to as gauge [13].

Mathematically, the gauge is delineated as follows:

$$\frac{\widehat{p}_{irr}}{p_{irr}},$$

$$E\left(\frac{\widehat{p}_{irr}}{p_{irr}}\right) \longrightarrow \alpha. \tag{5}$$

The gauge indicates the irrelevance part that corresponds to the nominal significance level (α), where p_{irr} shows a set of irrelevant covariates in the initial model and $\widehat{p}_{irr}$ shows the set of estimated irrelevant covariates [36].

Potency is defined as follows:

$$\frac{\widehat{p}_{rel}}{p_{rel}},$$

$$E\left(\frac{\widehat{p}_{irr}}{p_{irr}}\right) \longrightarrow 1. \tag{6}$$

This indicates that the relevant part p_{rel} shows the set of relevant covariates in the initial model and $\widehat{p}_{rel}$ points to the set of estimated relevant covariates, so the expected potency tending towards the value 1 is evidence of a good model [36]. Furthermore, we repeat each simulation experiment 1,000 times, and the expected potency and gauge evaluate the best method relatively. We use R software for the entire analysis.

3. Simulation Results and Discussion

The results of Monte Carlo experiments are illustrated in Tables 1–3.

> Scenario I: Table 1 presents simulation results for exponentially distributed errors, with varying sample sizes and covariates. All methods are improving with increasing sample size. In the case of low multicollinearity, in almost all cases, Autometrics and shrinkage methods such as SCAD and MCP hold all the relevant predictors, but shrinkage methods also hold irrelevant predictors in a large amount. Retaining irrelevant variables often lead to an overspecified model. Increasing the level of multicollinearity, Autometrics found 61% relevant variables (potency) along with around 3% irrelevant variables (gauge), while shrinkage methods retained more than 80% relevant variables with a much higher percentage of irrelevant variables. As we increase the sample size, the potency of Autometrics is dramatically enhanced and also gains improvement in gauge. Shrinkage methods improved the gauge, but it is still very high.
>
> Scenario II: Table 2 presents the simulation results for gamma distributed errors, with varying sample sizes and covariates. All results are improving with expanding the data window. In this scenario, all

Table 1: Simulation results for exponentially distributed errors.

{n, p}	Methods	0.25		0.90	
		Potency	Gauge	Potency	Gauge
{100, 20}	Autometrics	0.999	0.011	0.612	0.042
	SCAD	1	0.219	0.924	0.577
	MCP	1	0.140	0.925	0.562
{100, 25}	Autometrics	0.999	0.01	0.610	0.037
	SCAD	1	0.177	0.890	0.498
	MCP	1	0.110	0.885	0.472
{100, 30}	Autometrics	0.999	0.012	0.612	0.037
	SCAD	1	0.142	0.845	0.391
	MCP	1	0.085	0.847	0.397
{300, 20}	Autometrics	1	0.009	0.955	0.010
	SCAD	1	0.102	0.995	0.305
	MCP	1	0.090	0.995	0.294
{300, 25}	Autometrics	1	0.010	0.959	0.012
	SCAD	1	0.061	0.994	0.269
	MCP	1	0.056	0.992	0.261
{300, 30}	Autometrics	1	0.011	0.958	0.012
	SCAD	1	0.045	0.989	0.231
	MCP	1	0.039	0.990	0.230

Table 2: Simulation results for gamma distributed errors.

{n, p}	Methods	0.25		0.90	
		Potency	Gauge	Potency	Gauge
{100, 20}	Autometrics	1	0.01	1	0.014
	SCAD	1	0.101	0.999	0.065
	MCP	1	0.081	1	0.07
{100, 25}	Autometrics	1	0.012	1	0.016
	SCAD	1	0.070	1	0
	MCP	1	0.052	1	0.050
{100, 30}	Autometrics	1	0.013	1	0.022
	SCAD	1	0.047	1	0.046
	MCP	1	0.033	1	0.039
{300, 20}	Autometrics	1	0.01	1	0.013
	SCAD	1	0.100	1	0.027
	MCP	1	0.079	1	0.035
{300, 25}	Autometrics	1	0.011	1	0.012
	SCAD	1	0.070	1	0.031
	MCP	1	0.048	0.999	0.026
{300, 30}	Autometrics	1	0.010	1	0.009
	SCAD	1	0.055	1	0.026
	MCP	1	0.043	1	0.023

methods have correctly specified the relevant variables in most cases, but shrinkage methods have retained some irrelevant variables. In other words, it can be concluded that shrinkage methods overspecified the model.

Scenario III: Table 3 depicts the simulation findings for Frechet distributed errors, with varying sample sizes and covariates. The potency and gauge of almost all methods are improving with increasing sample size. In presence of low multicollinearity, all methods selected 100% relevant variables under a large sample. Autometrics has often selected around 1% irrelevant variables (gauge), while shrinkage methods have selected a

TABLE 3: Simulation results for Frechet distributed errors.

$\{n, p\}$	Methods	0.25		0.90	
		Potency	Gauge	Potency	Gauge
	Autometrics	0.995	0.01	0.719	0.032
$\{100, 20\}$	SCAD	0.999	0.155	0.965	0.499
	MCP	0.999	0.12	0.964	0.498
	Autometrics	0.996	0.011	0.731	0.031
$\{100, 25\}$	SCAD	0.999	0.115	0.937	0.438
	MCP	0.999	0.079	0.94	0.441
	Autometrics	0.995	0.011	0.734	0.050
$\{100, 30\}$	SCAD	1	0.089	0.918	0.374
	MCP	1	0.054	0.916	0.379
	Autometrics	0.999	0.01	0.983	0.009
$\{300, 20\}$	SCAD	1	0.091	0.999	0.179
	MCP	1	0.078	0.998	0.18
	Autometrics	1	0.010	0.987	0.010
$\{300, 25\}$	SCAD	1	0.063	0.997	0.17
	MCP	1	0.05	0.997	0.168
	Autometrics	1	0.011	0.985	0.012
$\{300, 30\}$	SCAD	1	0.045	0.995	0.153
	MCP	1	0.032	0.995	0.141

TABLE 4: Variables description.

Sr. No	Variables	Symbols	Source of data
1	Gold prices	GP	Yahoo! Finance
2	Gross domestic growth	GDP	SBP
3	Inflation	INF	WDI
4	Interest rate	IR	WDI
5	Unemployment	UEMP	WDI
6	Trade openness	TO	WDI
7	Stock market performance	SP	IFS
8	Real effective exchange rate	REER	WDI
9	Government stability	GS	ICRG
10	Credit premium risk	CPR	IFS
11	Oil prices	OP	Yahoo! Finance
12	Market rate	MR	IFS

Note: WDI stands for world development indicators; SBP stands for state bank of Pakistan; IFS stands for international financial statistics; and ICRG stands for international country risk guide.

large proportion of irrelevant variables. Increasing the level of multicollinearity, all methods are adversely affected. Autometrics retained 72% relevant variables with approximately 3% irrelevant variables. On the other hand, shrinkage methods hold more than 90% active variables along with a massive set of irrelevant variables. As we increase the number of observations, resultantly the potency of Autometrics improved and reduced the gauge to 1%. The improvement was achieved in the gauge of shrinkage methods, but it is still considered high.

Now we are comparing the potency and gauge of all methods across various error distributions. We can see that under the gamma distributed errors, the potency is higher and gauge is lower than the potency and gauge what we achieved under the exponentially and Frechet distributed errors.

4. Empirical Analysis

Complementing the Monte Carlo experiments, this study performs real data analysis using Pakistan financial data set. The data set consists of 12 time series observed at annual frequency spanning from 1981 to 2020 and is taken from the world development indicators, international financial statistics, Yahoo! Finance website, and international country risk guide. Among 12 variables, gold prices are the response variable, and the remaining variables are treated as predictors in this study. The predictors are selected through theories and literature to make a general model known as a general unrestricted model (GUM). Before analysis, some missing observations in the data set are replaced by averaging the neighbor observations and then standardizing the data set in order to reduce variation, which in turn provide stable results [37]. Detail regarding the variables has been

given in Table 4. Table 4 describes the variables, symbols, and source of data.

From Figure 2, it can be observed that the frequency distribution of the target variable (in our case, gold prices) is right-skewed, and the boxplot in Figure 2(b) also reveals that there are some outlying observations present in the series. However, Gujarati et al. [4] considered graphical representation as an informal approach therefore to reconfirm the distribution of gold prices; we move towards a statistical test that is known as the Shapiro test.

After applying the Shapiro test, we get a p-value that is almost zero; then the null hypothesis that the data are normally distributed is rejected. It implies that the distribution under consideration is highly skewed. Table 5 depicts the findings of real data considering 11 covariates. Autometrics hold GDP, IR, UEMP, TO, SP, and REER, which reveals that these covariates significantly contribute to gold prices. MCP selected all covariates except inflation (INF) and market rate (MR), and SCAD holds all covariates.

This is the fact that we do not know about the data-generating process in the real world. Therefore, it is difficult to compare the models' performance based on potency and gauge using real data. In such circumstances, the best and a widely used alternative approach is an out-of-sample forecast for models' assessment. But it requires dividing the data into two parts: a training set and a testing set. Thus, in this work, we split the data set into two parts: data from 1981 to 2010 are utilized to train the model, and the remaining data (2011–2020) are used to evaluate their forecasting performance. Root-mean-square error (RMSE) and mean absolute error (MAE) are computed to evaluate the forecasting performance of all considered methods, shown in Figure 3. The smaller the values of RMSE and MAE, the closer the predicted values to the actual values and resultantly indicates better forecast. The forecast errors were shown by the bar in Figure 3, which recommend that the Autometrics method outperformed the rival methods in the out-of-sample forecast. It illustrates that Autometrics has good predictive power than other competitor models in the sense that it is having the lowest prediction errors in multistep ahead forecast.
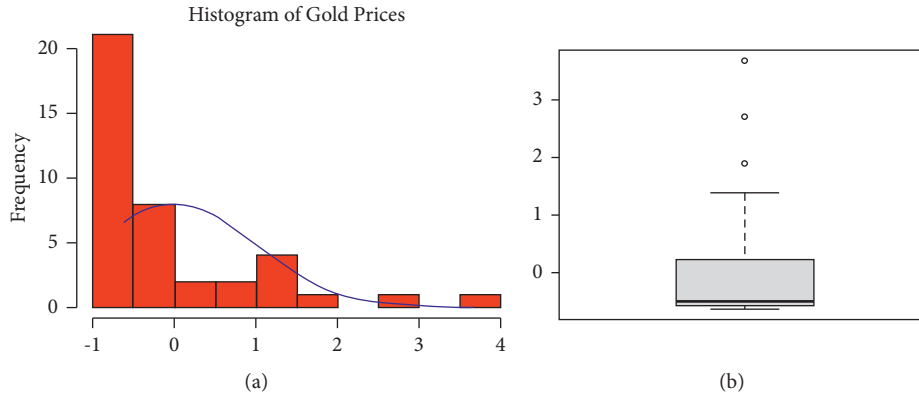
Figure 2: Histogram and boxplot of gold prices.

Table 5: Features selection based on real data.

Variables	MCP	Autometrics	SCAD
GDP	S	S	S
INF	NS	NS	S
IR	S	S	S
UEMP	S	S	S
TO	S	S	S
SP	S	S	S
REER	S	S	S
GS	S	NS	S
CPR	S	NS	S
OP	S	NS	S
MR	NS	NS	S

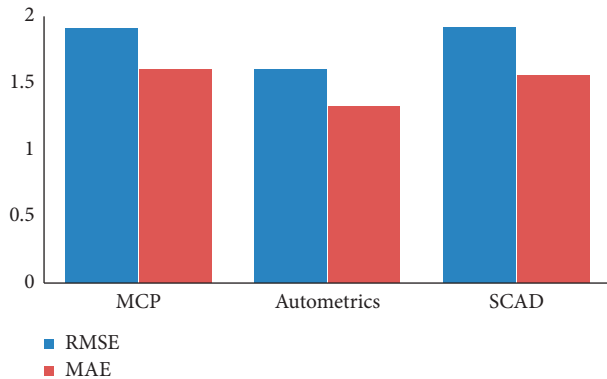Note: "S" shows the selected variables, and "NS" shows the nonselected variables.

Figure 3: Forecast comparison.

5. Conclusion Remarks

In this work, we compare Autometrics with two penalization techniques, that is, minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) under asymmetric error distributions such as exponential, gamma, and Frechet with altering sample sizes as well as predictors. Simulations using a wide variety of scenarios demonstrate that all methods improve for a large sample size. In the case of low multicollinearity, these methods perform well in terms of potency, but in terms of gauge, the shrinkage

methods collapse. Higher gauge leads to overspecification of the model. The increased level of multicollinearity among regressors adversely affects the performance of Autometrics and sparingly the shrinkage methods in terms of potency. At the same time, shrinkage methods select a massive set of irrelevant variables. We have observed that expanding the data window alleviates the detrimental influence of high multicollinearity on potency associated with Autometrics rapidly and steadily rectifies the gauge of penalized techniques.

For real data analysis, we consider the gold prices data along 11 covariates spanning from 1981 to 2020. To compare the forecasting performance of the selected methods, we divide the data into two parts, that is, 1981–2010 as training data and 2011–2020 as testing data. These methods are trained on training data, and their performance is assessed via testing data. Based on RMSE and MAE, Autometrics remained best in handling the gold prices trend and providing better forecasts than MCP and SCAD. We observed that penalization techniques hold many irrelevant covariates in comparison with Autometrics and hence tend to increase the forecast error comparatively.

Data Availability

Data can be provided upon special request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Zou and H. H. Zhang, "On the Adaptive Elastic-net with a diverging number of parameters," *Annals of Statistics*, vol. 37, no. 4, pp. 1733–1751, 2009.

[2] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, vol. 112, p. 18, New York, NY, USA, 2013.

[3] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[4] D. N. Gujarati, D. C. Porter, and S. Gunasekar, *Basic Econometrics*, Tata McGraw-Hill Education, New York, NY, USA, 2012.

[5] S. Ali, H. Khan, I. Shah, M. M. Butt, and M. Suhail, "A comparison of some new and old robust ridge regression estimators," *Communications in Statistics - Simulation and Computation*, vol. 50, no. 8, pp. 2213–2231, 2021.

[6] J. Inglis and E. E. Leamer, "Specification searches: ad hoc inference with nonexperimental data," *Technometrics*, vol. 23, no. 1, p. 112, 1981.

[7] H. R. Varian, "Big data: new tricks for econometrics," *The Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014.

[8] N. R. Swanson and W. Xiong, "Big data analytics in economics: what have we learned so far, and where should we go from here?" *Canadian Journal of Economics/Revue canadienne d'économique*, vol. 51, no. 3, pp. 695–746, 2018.

[9] J. L. Castle, J. A. Doornik, and D. F. Hendry, "Modelling non-stationary "Big data"," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1556–1575, 2020.

[10] N. R. Ericsson, *Detecting Crises, Jumps, and Changes in Regime*, Board of Governors of the Federal Reserve System, Washington, DC, USA, 2012.

[11] J. L. Castle, M. P. Clements, and D. F. Hendry, "Forecasting by factors, by variables, by both or neither?" *Journal of Econometrics*, vol. 177, no. 2, pp. 305–319, 2013.

[12] J. Castle, J. Doornik, D. Hendry, and F. Pretis, "Detecting location shifts during model selection by step-indicator saturation," *Econometrics*, vol. 3, no. 2, pp. 240–264, 2015.

[13] J. A. Doornik and D. F. Hendry, "Statistical model selection with "big data"," *Cogent Economics & Finance*, vol. 3, no. 1, Article ID 1045216, 2015.

[14] F. Pretis, L. Schneider, J. E. Smerdon, and D. F. Hendry, "Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation," *Journal of Economic Surveys*, vol. 30, no. 3, pp. 403–429, 2016.

[15] C. Epprecht, D. Guégan, Á. Veiga, and J. Correa da Rosa, "Variable selection and forecasting via automated methods for linear models: LASSO/adaLASSO and Autometrics," *Communications in Statistics - Simulation and Computation*, vol. 50, no. 1, pp. 103–122, 2021.

[16] J. V. Rocha and P. L. V. Pereira, *The Effects of Estimation Sample Size in Forecast Performance: The Case of Brazilian Industrial Production Index*, UDES, South Region, Brazil, 2019.

[17] C. De Mol, D. Giannone, and L. Reichlin, "Forecasting using a large number of predictors: is bayesian regression a valid alternative to principal components?" *SSRN Electronic Journal*, vol. 146, no. 2, pp. 318–328, 2006.

[18] A. Inoue and L. Kilian, "How useful is bagging in forecasting economic time series? a case study of U.S. Consumer price inflation," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 511–522, 2008.

[19] J. Bai and S. Ng, "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, vol. 146, no. 2, pp. 304–317, 2008.

[20] H. H. Kim and N. R. Swanson, "Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence," *Journal of Econometrics*, vol. 178, pp. 352–367, 2014.

[21] N. R. Swanson, "Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods," *International Journal of Forecasting*, vol. 34, no. 2, pp. 339–354, 2018.

[22] M. Luciani, "Forecasting with approximate dynamic factor models: the role of non-pervasive shocks," *International Journal of Forecasting*, vol. 30, no. 1, pp. 20–29, 2014.

[23] N. R. Swanson and W. Xiong, *Predicting Interest Rates Using Shrinkage Methods, Real-Time Diffusion Indexes, and Model Combinations*, Rutgers University, Newark, NJ, USA, 2018.

[24] N. R. Swanson, W. Xiong, and X. Yang, "Predicting interest rates using shrinkage methods, real-time diffusion indexes, and model combinations," *Journal of Applied Econometrics*, vol. 35, no. 5, pp. 587–613, 2020.

[25] K. Maehashi and M. Shintani, "Macroeconomic forecasting using factor models and machine learning: an application to Japan," *Journal of the Japanese and International Economies*, vol. 58, p. 101104, 2020.

[26] D. F. Hendry and H. M. Krolzig, "The p," *The Economic Journal*, vol. 115, no. 502, pp. C32–C61, 2005.

[27] J. A. Doornik, *Econometric Model Selection with More Variables than Observations*, Economics Department, University of Oxford, Oxford, UK, 2009.

[28] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete l_1 linear approximation," *SIAM Journal on Numerical Analysis*, vol. 10, no. 5, pp. 839–848, 1973.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[30] Z. Y. Algamal and M. H. Lee, "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification," *Computers in Biology and Medicine*, vol. 67, pp. 136–145, 2015.

[31] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[32] M. Lu, J. Zhou, C. Naylor et al., "Application of penalized linear regression methods to the selection of environmental enteropathy biomarkers," *Biomarker research*, vol. 5, no. 1, pp. 1–10, 2017.

[33] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[34] A. Wahid, D. M. Khan, and I. Hussain, "Robust Adaptive Lasso method for parameter's estimation and variable selection in high-dimensional sparse models," *PLoS one*, vol. 12, no. 8, Article ID e0183518, 2017.

[35] Z. Ahmad, E. Mahmoudi, and O. Kharazmi, "On modeling the earthquake insurance data via a new member of the T-X family," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 7631495, 20 pages, 2020.

[36] F. Pretis, J. Reade, and G. Sucarrat, "Automated General-to-Specific (GETS) regression modeling and indicator saturation methods for the detection of outliers and structural breaks," *Journal of Statistical Software*, vol. 86, no. 3, 2018.

[37] I. Tsamardinos, G. Borboudakis, P. Katsogridakis, P. Pratikakis, and V. Christophides, "A greedy feature selection algorithm for big data of high dimensionality," *Machine Learning*, vol. 108, no. 2, pp. 149–202, 2019.