



HAL
open science

Semi-Independent Sliced Inverse Regression

Kevin Li

► **To cite this version:**

| Kevin Li. Semi-Independent Sliced Inverse Regression. 2014. hal-00960343

HAL Id: hal-00960343

<https://hal.science/hal-00960343>

Preprint submitted on 18 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-Independent Sliced Inverse Regression

Kevin B. Li

Beijing Jiaotong University

Abstract

This paper deals with dimension reduction in regression for large dataset. A new method based on the sliced inverse regression approach is introduced, called semi-independent regularized sliced inverse regression. Our method can handle highly correlated variables. Asymptotic properties are established under some linearity conditions. An application on an economic dataset shows that our method outperforms the non-stationary factor model.

1 Introduction

Forecasting using many predictors has received a good deal of attention in recent years. The curse of dimensionality has been turned into a blessing with the abundant information in large datasets. Various methods have been originated to extract efficient predictors, for example, non-stationary factor model (NFM), Bayesian model averaging, Lasso, boosting, etc. Among them, non-stationary factor model is conceptually appealing in economics because it is structurally consistent with log-linearized models such as non-stationary stochastic general equilibrium models. The recent development in statistics provides a new method of dimension reduction in regression for large-dimensional data. The literature stems from [24], and [37], which proposed a new way of thinking in the regression analysis, called sliced inverse regression (SIR). SIR reverses the role of response y and predictors \mathbf{x} . Classical regression methods mainly deal with the conditional density $\mathbf{f}(y|\mathbf{x})$. SIR collects the information of the variation of predictors \mathbf{x} along with the change of the response y , by exploring the conditional density $h(\mathbf{x}|y)$. Usually the

dimension of the response is far more less than the dimension of the predictors, hence, it is a way to avoid the “curse of dimensionality”. The traditional SIR does not work well for highly correlated data, because the algorithm requires the inverse of the covariance matrix. This is not feasible when the number of variables N is greater than the number of observations T , which happens a lot in economics studies. In addition, the economic variables are often highly correlated, due to the derivation formula or same category. This makes the covariance matrix ill-conditioned, causes the inverse matrix lack of precision and too sensitive to the variation of matrix entries, and leads to a false or unstable result. There are some extensions of SIR for the highly collinearity data and “ $T < N$ ” problems, for example, regularized sliced inverse regression ([76], [44]) and partial inverse regression ([42]). In this paper, we propose a new method of dimension reduction, called the semi-independent sliced inverse regression (SISIR) method, for many predictors in a data rich environment. We evaluate its property theoretically and use it for forecasting economic series. Comparison in terms of pseudo out-of-sample forecasting simulation shows the advantage of our method. The remaining of the paper is organized as follows. Section 2 introduces semi-independent SIR method with its statistical property. An application on an economic dataset is given in Section 3. Conclusions with some discussions are provided in Section 4.

2 Modeling and methods

The regression model in [37] takes the form of

$$y = g(\beta_1' \mathbf{x}, \beta_2' \mathbf{x}, \dots, \beta_K' \mathbf{x}, \epsilon), \quad (2.1)$$

where the response y is univariate, \mathbf{x} is an N -dimensional vector, and the random error ϵ is independent of \mathbf{x} . Many methods can be used to find the e.d.r.-directions, for example, principal component analysis might be the most commonly used one in economics. But unlike these methods, SIR not only reduces dimensions in regression but also integrates the information from both predictors and response. Moreover, different from the classical regression methods, SIR intends to collect information on how \mathbf{x} changes along with y . That is to say, instead of estimating the forward regression function $\boldsymbol{\eta}(\mathbf{x}) = E(y|\mathbf{x})$, inverse regression considers $\boldsymbol{\xi}(y) = E(\mathbf{x}|y)$. Compared with $\boldsymbol{\eta}(\mathbf{x})$, the inverse regression function $\boldsymbol{\xi}(y)$ depends on one-dimensional y ,

which makes the operation much easier. [37] showed that using SIR method, the e.d.r.-directions can be estimated by solving

$$\text{Cov}(E(\mathbf{x}|y))\boldsymbol{\beta}_j = \nu_j \text{Cov}(\mathbf{x})\boldsymbol{\beta}_j, \quad (2.2)$$

where ν_j is the j th eigenvalue and $\boldsymbol{\beta}_j$ is the corresponding eigenvector of $\text{Cov}(E(\mathbf{x}|y))$ with respect to $\text{Cov}(\mathbf{x})$. During the forecasting procedure, the covariance matrices can be replaced by their usual moment estimates.

2.1 Semi-independent sliced inverse regression

In this section, we introduce grouping methodology with the sliced inverse regression to improve the performance of SIR on collinear data. Assume that the variables of interest can be grouped into several blocks, so that two variables within the same block are correlated to each other, and any two variables belonging to different blocks are independent. In practice, an orthogonalization procedure can be applied to reduce the correlations between blocks in order to fit our assumption. Thus, we can group the variables according to their correlations in order to find the e.d.r.-directions, because there is no shared information between groups. The grouping method we use is hierarchical grouping with complete linkage. The dissimilarity is defined as $1 - |\text{Correlation}|$. The algorithm for the semi-independent SIR method can be described as following.

1. Standardize each explanatory variable to zero mean and unit variance.
2. Cluster \mathbf{x} ($N \times 1$) into $(\mathbf{x}_1' \ \cdots \ \mathbf{x}_c')'$ based on the correlation matrix of \mathbf{x} , where \mathbf{x}_i is $N_i \times 1$, $\sum_{i=1}^c N_i = N$, and c is the number of groups, which will be determined by cross-validation.
3. Restricted to each group, perform SIR method and pick up k_i SIR directions based on the sequential chi-square test (Li, 1991), say $\boldsymbol{\theta}_j^{(i)}$, $j = 1, \dots, k_i$, $i = 1, \dots, c$.
4. Collect all the SIR variates obtained from the groups, say $\{\boldsymbol{\theta}_j^{(i)'} \mathbf{x}_i, i = 1, 2, \dots, c, j = 1, 2, \dots, k_i\}$.
5. Let $\boldsymbol{\lambda}_l = (\mathbf{0}_1' \ \boldsymbol{\theta}_j^{(i)'} \ \mathbf{0}_2')'$, $l = 1, \dots, m$, $m = \sum_{i=1}^c k_i$, where $\mathbf{0}_1$ and $\mathbf{0}_2$ are zero column vectors with dimension $\sum_{k=1}^{i-1} N_k$ and $\sum_{k=i+1}^c N_k$, respectively. Denote $\Lambda = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_m)$. The variates $\{\boldsymbol{\theta}_j^{(i)'} \mathbf{x}_i\}$ can be written in a vector form as $(\boldsymbol{\lambda}'_1 \mathbf{x}, \dots, \boldsymbol{\lambda}'_m \mathbf{x})' = \Lambda' \mathbf{x}$.

6. Then, perform SIR method one more time to the pooled variates $\Lambda'\mathbf{x}$ to reduce dimensions further, and get the e.d.r.-directions $(\gamma_1, \gamma_2, \dots, \gamma_v)$, where v is also determined by the sequential chi-square test. Denote $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_v)$, the final SISIR variates we chose are $\Gamma'\Lambda'\mathbf{x}$.
7. Estimate the values of forecasting series using the SISIR variates $\Gamma'\Lambda'\mathbf{x}$. Linear models are used in this paper.

Note that the matrices Γ is $m \times v$, Λ is $N \times m$, so $\Gamma'\Lambda'\mathbf{x}$ is $v \times 1$. Therefore, we only use v factors to build the final model for forecasting y , instead of using N variables based on the original dataset.

2.2 Statistical property of semi-independent SIR

[37] established the unbiasedness for the e.d.r.-directions found by SIR, assuming the following linearity condition.

Linearity Condition 1. For any $\mathbf{b} \in \mathbb{R}^N$, the conditional expectation $E(\mathbf{b}'\mathbf{x}|\beta'_1\mathbf{x}, \dots, \beta'_K\mathbf{x})$ is linear in $\beta'_1\mathbf{x}, \dots, \beta'_K\mathbf{x}$.

The linearity condition is not easy to verify, however, [26] showed when \mathbf{x} is elliptically symmetrically distributed, for example, multivariate normally distributed, the linearity condition holds. Furthermore, [32] showed that elliptical symmetric distribution is not a restrictive assumption, because the linearity condition holds approximately when N is large even if the dataset is not elliptically symmetric. Without loss of generality, we assume each variable in \mathbf{x} has been standardized to zero mean and unit variance for our discussion. [37] proved the following theorem,

Theorem 1. *Assume Linearity Condition 1, the centered inverse regression curve $E(\mathbf{x}|y)$ is contained in the space spanned by $\Sigma_{\mathbf{x}}\beta_j$, $j = 1, \dots, K$, where $\Sigma_{\mathbf{x}}$ is the covariance matrix of \mathbf{x} .*

Similar unbiasedness property can be proved for our semi-independent SIR.

Theorem 2. *Under certain linearity conditions, $E(\mathbf{x}|y)$ is contained in the space spanned by $\Sigma_{\mathbf{x}}\Lambda\Gamma$.*

Theorem 2 describes the desirable property that there is no estimation bias. The e.d.r.-space estimated by our SISIR method contains the true inverse regression curve.

2.3 Orthogonalization

For a given dataset \mathbf{X} with dimension $N \times T$, and groups $\mathbf{X}_1, \dots, \mathbf{X}_c$, the correlations between these groups need to be reduced to zero, to achieve group-wise independence. QR decomposition along with projection operators is used to perform the orthogonalization. To begin with, use QR decomposition to find the orthogonal bases of the first group \mathbf{X}_1 , named as \mathbf{Q}_1 . Next, project the second group \mathbf{X}_2 onto the space of $\text{span}\{\mathbf{Q}_1\}^\perp$, which is the orthogonal complement of the space spanned by \mathbf{X}_1 , named as \mathbf{X}_2^* ,

$$\mathbf{X}_2^* = (\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1')\mathbf{X}_2. \quad (2.3)$$

Then use QR decomposition again to find the orthogonal bases of \mathbf{X}_2^* , named as \mathbf{Q}_2 , and project \mathbf{X}_3 onto the space of $\text{span}\{\mathbf{Q}_1, \mathbf{Q}_2\}^\perp$, named as \mathbf{X}_3^* . Keep doing such process till the last group \mathbf{X}_c , we will get a new sequence of groups $\mathbf{X}_1, \mathbf{X}_2^*, \dots, \mathbf{X}_c^*$, in which every two groups are orthogonal, and the new sequence contains all the information of the original dataset \mathbf{X} .

2.4 Regularization

Due to the high correlations between the series within each group, the covariance matrices of each group $\Sigma_{\mathbf{x}_i}$ are ill-conditioned, which make them hard to be inverted. We suggest a regularized version of the covariance matrix to overcome this issue ([28]).

$$\Sigma_{\mathbf{x}_i}(\tau) = (1 - \tau)\Sigma_{\mathbf{x}_i} + \tau \frac{\text{tr}\Sigma_{\mathbf{x}_i}}{N_i} I_{N_i}, \quad (2.4)$$

where $\tau \in [0, 1]$ is the shrinkage parameter. This is similar to the ridge version proposed by [76], which replaces $\Sigma_{\mathbf{x}_i}$ with $\Sigma_{\mathbf{x}_i} + \tau I_{N_i}$. The shrinkage parameter τ can be chosen by cross-validation. Note when $\tau = 1$, the regularized covariance matrix will degenerate to a diagonal matrix whose diagonal elements are the means of the eigenvalues of $\Sigma_{\mathbf{x}_i}$. In such case, the chosen e.d.r.-direction is one of the input series, and the other series, which may also contain information for the predictors, are discarded.

2.5 Comparison between SISIR and SIR

Before applying the proposed SISIR method to real data, consider the following simulated example first to compare the performance of SISIR and SIR

methods. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{10}$ be independent and identically distributed (i.i.d.) with multivariate normal distribution $N(0, \Sigma)$, where Σ is a 10×10 covariance matrix with 1 at diagonal and 0.9 at off-diagonal. The random error \mathbf{e} is independent to \mathbf{x}_i 's, and follows normal distribution $N(0, 0.1)$. The response y is simulated using the following formula,

$$y = \sum_{j=1}^{10} j \times \mathbf{x}_j + \mathbf{e}$$

Root mean square error (RMSE) is considered as a criterion to evaluate the prediction performance.

$$\text{RMSE} = \sqrt{\sum_{i=1}^T (\hat{y}_i - y_i)^2 / T}, \quad (2.5)$$

where \hat{y}_i is the i th predicted value of response, y_i is the i th observed value, and T is the number of observations. We simulate 300 observations at each run under above conditions. In SISIR, the parameters c and τ are chosen as $c = 10$, $\tau = 0.5$ to minimize RMSE. Table 1 presents the means and standard deviations for the RMSE of SISIR and SIR across 100 runs. From Table 1,

Table 1: RMSE for SISIR and SIR

	SISIR	SIR
Mean	11.73	17.04
S.D.	0.81	1.95

it is clear that SISIR has much smaller RMSE than SIR. In fact, our other simulations, which are not presented here, show that SISIR performs even better when the sample size T decreases to N .

3 Application

The dataset contains 143 quarterly economic variables from 13 economic categories. We use 109 of them as explanatory variables, since the other 34 are just high-level aggregates of the 109. All 143 variables are used for

forecasting purpose. The correlation plot of the 109 predictor series after logarithm and/or differencing transformation is showed in Figure 1, which demonstrates that there do exist some highly correlated blocks. Therefore, our semi-independent method is necessary for this dataset. For the purpose of comparison, similar rolling pseudo out-of-sample forecasting simulation is used. The main steps can be described as follows,

1. From 1985 to 2008, at each date t , use cross-validation, which is described below, to the most recent 100 observations to choose the parameter c and τ in SISIR based on mean square error.
2. Use the chosen \hat{c} and $\hat{\tau}$, apply SISIR one more time to predict \hat{y}_{t+h} , where h is the forecasting period.
3. Calculate the RMSE for the forecasting procedure,

$$\text{RMSE} = \sqrt{\sum_{t=1}^T (y_{t+h} - \hat{y}_{t+h})^2 / T}.$$

The steps for cross-validation are described as follows,

1. Regress y_{t+h} and x_t on the autoregressive terms 1, y_t , y_{t-1} , y_{t-2} , y_{t-3} , in order to eliminate the autoregressive effect. Denote the residuals as \tilde{y}_{t+h} and \tilde{x}_t .
2. Let $\mathfrak{S}(t) = \{1, \dots, t - 2h - 3, t + 2h + 3, \dots, 100\}$, at each date $t = 1, \dots, 100 - h$, find the e.d.r-directions and linear regression model using SISIR and observations \tilde{y}_i and \tilde{x}_i , $i \in \mathfrak{S}(t)$.
3. Use the e.d.r-directions and linear regression model from the above step at date t to predict \tilde{y}_{t+h} .
4. For fixed h , parameters (c, τ) are chosen by minimizing the sum of squared forecasting error,

$$(\hat{c}, \hat{\tau}) = \operatorname{argmin} \frac{1}{100 - h} \sum_{t=1}^{100-h} \left(\tilde{y}_{t+h} - \hat{\tilde{y}}_{t+h} \right)^2.$$

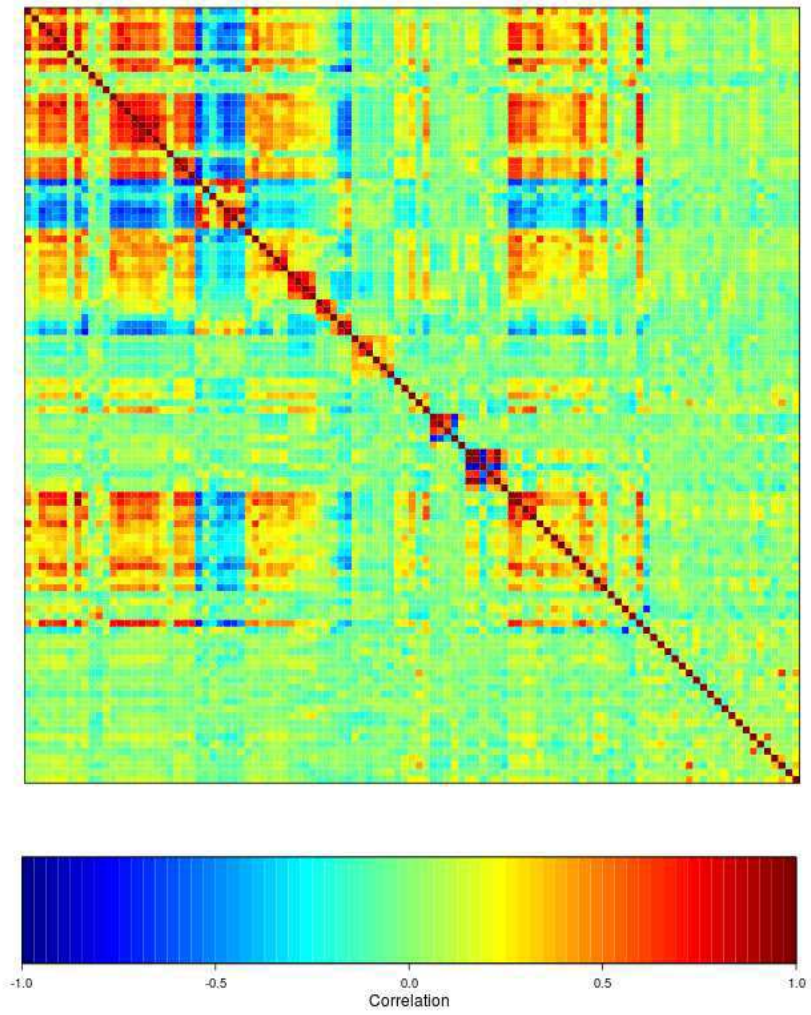


Figure 1: Plot of Correlations of the 109 Predictor Series

We compare our method with the non-stationary factor model using the first five principle components (NFM-5), which was claimed to be no worse than any other shrinkage methods according to Stock and Watson (2011). Autoregressive model of order 4 (AR(4)) is used as a benchmark, all RMSEs are recorded as the ratio relative to AR(4). Smaller relative RMSE indicates better forecasting performance. Table 2 presents the number of series with smaller RMSE than AR(4) model for SISIR and NFM-5. We can see that for forecasting period $h = 1$, if SISIR is used, there are 97 series out of 143 have smaller RMSE than the benchmark AR(4) model. If NFM-5 is used, only 85 series out of 143 have smaller RMSE than AR(4) model. The differences become even larger for big forecasting period, when $h = 4$ the number of series of SISIR increases to 115 while the number of NFM-5 decreases to 53. Table 3 presents the distributions of the RMSEs for AR(4), NFM-5, and

Table 2: Number of Series with Smaller RMSE than AR(4) Model

	NFM-5	SISIR
$h = 1$	85	97
$h = 2$	59	109
$h = 4$	53	115

SISIR methods. When $h = 1$, the first quartile of the relative RMSE of SISIR is just 0.768, which is much smaller than the relative RMSE of NFM-5 (0.961), and the median relative RMSE of SISIR is 0.907, while NFM-5 has 0.993. When $h = 2$ and $h = 4$, SISIR improves the forecasting results of AR(4) for more than 3/4 of the series. The relative RMSE of SISIR at first, second, and third quartile are all smaller than those of NFM-5. From Table 2 and 3, one can tell that SISIR improves the forecasting results significantly compared to the NFM-5 method, especially for longer forecasting period.

Table 4 presents the median RMSE relative to AR(4) model by category via cross-validation. Column “S&W” reports the smallest relative RMSE Stock and Watson got using NFM-5 and other shrinkage methods in their 2011 paper. Comparing all these results, SISIR method has smaller median relative RMSEs for more than 70% of these categories among three forecasting period, which demonstrates its superiority again. Table 4 also indicates the performance of SISIR varied across categories. It has outstanding performance for some categories, such as **Industrial Production**, **Unemployment**

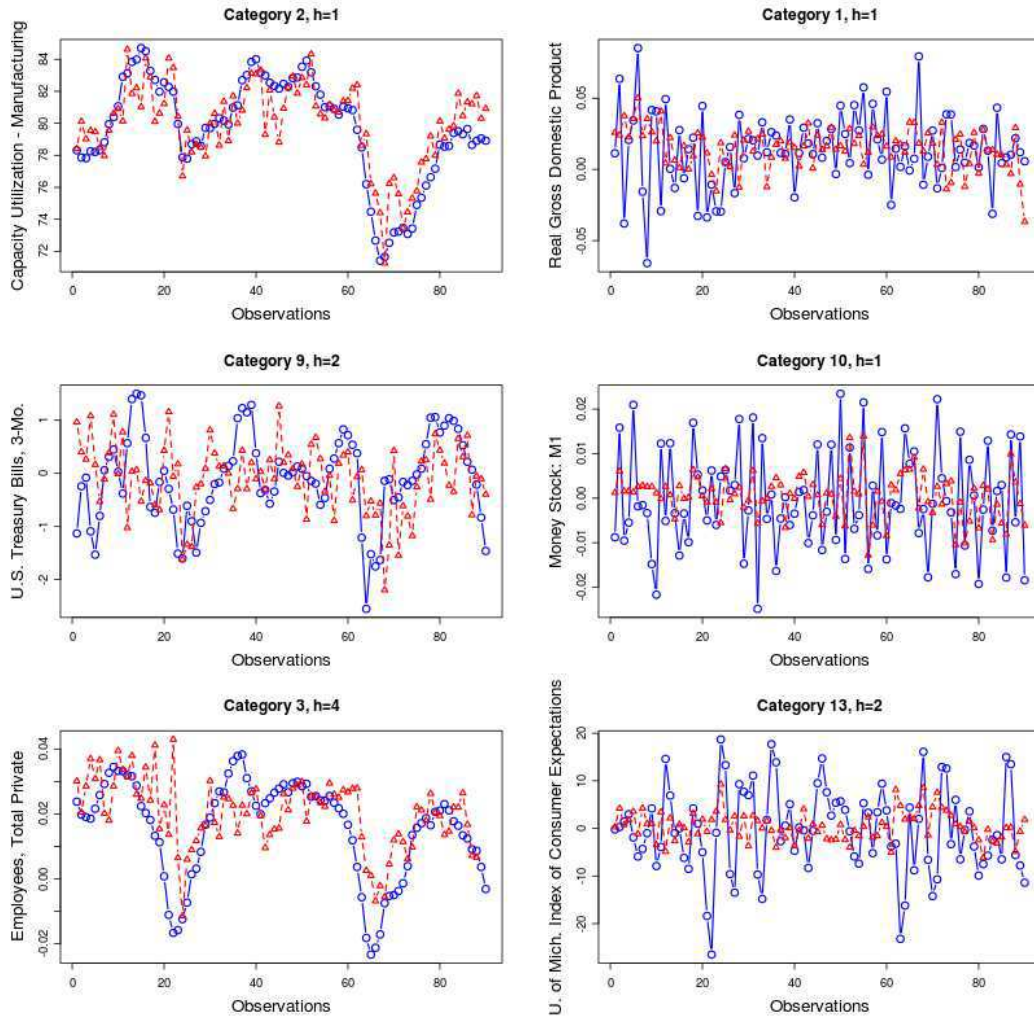
Table 3: Distributions of Relative RMSEs by Pseudo Out-of-Sample Forecasting

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	1.000	1.000	1.000	1.000	1.000
NFM-5	0.874	0.961	0.993	1.022	1.089
SISIR	0.621	0.768	0.907	1.048	1.372

(a) $h = 1$

Table 4: Median Relative RMSE for Forecasting by Category of Series

Category	$h = 1$			$h = 2$			$h = 4$		
	NFM-5	S&W	SISIR	NFM-5	S&W	SISIR	NFM-5	S&W	SISIR
1. GDP Components	0.905	0.905	1.079	0.907	0.870	0.807	0.906	0.906	0.839
2. Industrial Production	0.882	0.882	0.669	0.861	0.852	0.694	0.827	0.827	0.745
3. Employment	0.861	0.861	0.849	0.861	0.859	0.803	0.844	0.842	0.823
4. Unempl. Rate	0.800	0.799	0.771	0.750	0.723	0.723	0.762	0.743	0.647
5. Housing	0.936	0.897	1.220	0.940	0.902	1.081	0.926	0.882	0.807
6. Inventories	0.900	0.886	0.856	0.867	0.867	0.764	0.856	0.856	0.784
7. Prices	0.980	0.970	0.865	0.977	0.961	0.892	0.963	0.948	0.797
8. Wages	0.993	0.938	0.967	0.999	0.919	0.960	1.019	0.931	1.031
9. Interest Rates	0.980	0.946	0.849	0.952	0.928	0.892	0.956	0.949	0.822
10. Money	0.953	0.926	1.000	0.933	0.921	0.950	0.909	0.909	0.927
11. Exchange Rates	1.015	0.981	0.974	1.015	0.980	1.108	1.036	0.965	1.150
12. Stock Prices	0.983	0.983	0.840	0.977	0.955	0.893	0.974	0.961	1.039
13. Cons. Exp.	0.977	0.977	0.765	0.963	0.960	1.082	0.966	0.955	0.963



(a) From SISIR Favored Categories

(b) From SISIR No-Favored Categories

Figure 2: Plots of the Forecasting Values (Δ) vs. Real Observations (\circ) from 1985 to 2008.

Rate, Inventories, Interest Rates, etc. But it does not work well for some others, such as Housing, Money, Exchange Rates. Figure 2 plots six series from both SISIR favored and no-favored categories. Three of them in Figure 2a are from SISIR favored categories and three of them in Figure 2b are from SISIR no-favored categories. From these plots, one can see that the responses of the SISIR no-favored series are quite disordered. They are more like white noises, the variations are big but the changes of \mathbf{x} means are not distinct. The inverse regression method is aimed to detect the variation of $E(\mathbf{x}|y)$. If the conditional expectations of \mathbf{x} do not have much difference for different values of y , the estimation for the e.d.r.-directions will be inaccurate, and will lead to the poor performance on forecasting.

4 Conclusion

Sliced inverse regression now becomes a popular dimension reduction method in computer science, engineering and biology. In this article, we bring it to economic forecasting model when there is a large number of predictors and high collinearity. Compared to the classical non-stationary factor model, SIR retrieves information not only from the predictors but also from the response. Moreover, our semi-independent regularized SIR has the ability to handle highly collinearity or “ $T < N$ ” data. The simulation confirms that it offers a lot of improvements over NFM-5 model on the economic dataset. After finding the SISIR variates, we use linear models for forecasting the responses y , because scatter plots for SISIR variates and y values show strong linear relationships, and the results are desirable. But one may use polynomials, splines, Lasso, or some other more advanced regression techniques for different situations. Based on its basic idea, there are more than one generalizations of SIR using higher order inverse moments. For instance, SAVE, SIR-II ([37]), DR ([45]), and SIMR ([71]). Our semi-independent algorithm can also be applied to these methods for highly collinearity data, and good performance is expected. Above all, we can conclude that the semi-independent regularized sliced inverse regression is a powerful tool in forecasting using many predictors. It may not be limited in economic forecasting, and can also be applied to dimension reduction or variable selection problems in social science, microarray analysis, or clinical trails when the dataset is large and highly correlated.

References

- [1] Aragon, Y. (1997). A Gauss implementation of multivariate sliced inverse regression. *Computational Statistics*, 12, 355–372.
- [2] Aragon, Y. and Saracco, J. (1997). Sliced Inverse Regression (SIR): an appraisal of small sample alternatives to slicing. *Computational Statistics*, 12, 109–130.
- [3] Azaïs, R., Gégout-Petit, A., and Saracco, J. (2012). Optimal quantization applied to sliced inverse regression. *Journal of Statistical Planning and Inference*, 142, 481–492.
- [4] Barreda, L., Gannoun, A., and Saracco, J. (2007). Some extensions of multivariate SIR. *Journal of Statistical Computation and Simulation*, 77, 1–17.
- [5] Bentler, P.M., Xie, J., (2000). Corrections to test statistics in principal Hessian directions. *Statistics and Probability Letters*, 47, 381–389.
- [6] Bernard-Michel, C., Gardes, L. and Girard, S. (2008). A Note on Sliced Inverse Regression with Regularizations, *Biometrics*, 64, 982–986.
- [7] Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L. and Girard, S. (2009). Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression. *Journal of Geophysical Research - Planets*, 114, E06005.
- [8] Bernard-Michel, C., Gardes, L. and Girard, S. (2009). Gaussian Regularized Sliced Inverse Regression, *Statistics and Computing*, 19, 85–98.
- [9] Bura, E. and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society, Series B*, 63, 393–410.
- [10] Chavent, M., Kuentz, V., Lique, B., and Saracco, J. (2011). A sliced inverse regression approach for a stratified population. *Communications in statistics - Theory and methods*, 40, 1–22.
- [11] Chavent, M., Girard, S., Kuentz, V., Lique, B., Nguyen, T.M.N. and Saracco, J. (2014). A sliced inverse regression approach for data stream, *Computational Statistics*, to appear.

- [12] Chen, C.-H. and Li, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statistica Sinica*, 8, 289–316.
- [13] Cook, R.D., Weisberg, S., (1991). Discussion of ‘sliced inverse regression for dimension reduction’. *Journal of the American Statistical Association*, 86, 328-332.
- [14] Cook, R.D., (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. Proceedings of the Section on Physical and Engineering Sciences. Alexandria, VA: American Statistical Association. 18-25.
- [15] Cook, R.D., 1996. Graphics for regressions with a binary response. *Journal of the American Statistical Association*, 91, 983-992.
- [16] Cook, R. D.(1998). Principal hessian directions revisited (with discussion). *Journal of the American Statistical Association*, 93, 84–100.
- [17] Cook, R.D., (1998b). Regression Graphics, Ideas for Studying Regressions through Graphics. Wiley, New York.
- [18] Cook, R.D., Lee, H., (1999). Dimension-reduction in binary response regression. *Journal of the American Statistical Association*, 94, 1187-1200.
- [19] Cook, R. D. (2000). SAVE: a method for dimension reduction and graphics in regression. *Communications in statistics - Theory and methods*, 29, 2109–2121.
- [20] Cook, R.D., Critchley, F., (2000). Identifying regression outliers and mixtures graphically. *Journal of the American Statistical Association*, 95, 781-794.
- [21] Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics*, 30, 450–474.
- [22] Cook, R.D., Ni, L., (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, 100, 410-428.

- [23] Coudret, R., Girard, S. and Saracco, J. (2014). A new sliced inverse regression method for multivariate response regression, *Computational Statistics and Data Analysis*, to appear.
- [24] Duan, N. and Li, K.-C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, 19, 505–530.
- [25] Dunia, R. and Joe Qin, S. (1998). Subspace approach to multidimensional fault identification and reconstruction. *AIChE Journal*, 44, 1813–1831.
- [26] M.L. Eaton, A characterization of spherical distributions, *Journal of Multivariate Analysis*, 1986, 20(2):272–276.
- [27] Ferré, L. (1998). Determining the dimension in sliced inverse regression and related methods. *Journal of the American Statistical Association*, 93, 132–140.
- [28] J.H. Friedman, Regularized discriminant analysis, *Journal of the American statistical association*, 1989, 84(405):165–175.
- [29] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference ranges based on nonparametric quantile regression, *Statistics in Medicine*, 21, 3119-3135.
- [30] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2004). Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46, 103–122.
- [31] Gannoun, A. and Saracco, J. (2003). An asymptotic theory for SIR_α method. *Statistica Sinica*, 13, 297–310.
- [32] Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, 21, 867-889.
- [33] Hsing, T. (1999). Nearest neighbor inverse regression. *The Annals of Statistics*, 27, 697–731.
- [34] Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics*, 20, 1040–1061.

- [35] Kuentz, V., Liquet, B., and Saracco, J. (2010). Bagging versions of sliced inverse regression. *Communications in statistics - Theory and methods*, 39, 1985–1996.
- [36] Kuentz, V. and Saracco, J. (2010). Cluster-based sliced inverse regression. *Journal of the Korean Statistical Society*, 39, 251–267.
- [37] Li, K.C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 86, 316–342.
- [38] Li, K.C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Steins lemma. *Journal of the American Statistical Association*, 87, 1025–1039.
- [39] K.C. Li, High dimensional data analysis via the sir/phd approach, 2000.
- [40] Li, K.-C., Aragon, Y., Shedden, K., and Agnan., C. T. (2003). Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, 98, 99–109.
- [41] Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. *Technometrics*, 48, 503–510.
- [42] Li, L., Cook, R. D. and Tsai, C. L. (2007) Partial inverse regression. *Biometrika*, 94, 615–625.
- [43] Li, B., Wang, S., (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102, 997–1008.
- [44] Li, L. and Yin, X. (2008). Sliced inverse regression with regularizations, *Biometrics*, 64, 124–131.
- [45] Li, B. and Wang, S. (2007). On directional regression for dimension reduction, *Journal of the American Statistical Association*, 102, 997–1008.
- [46] Li, K.B. (2013a). Invariance properties of Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00805491>
- [47] Li, K.B. (2013b). A review on Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00803698>

- [48] Li, K.B. (2013c). Some limitations of Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00803602>
- [49] Li, K.B. (2013d). Standardized Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00833686>
- [50] Li, K.B. (2013e). Independent Sliced Inverse Regression, *Journal of Parametric and Non-Parametric Statistics*, **1**, <http://jpanps.altervista.org>
- [51] Liqueet, B. and Saracco, J. (2008). Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR_α method. *Communications in statistics - Simulation and Computation*, *37*, 1198–1218.
- [52] Liqueet, B. and Saracco, J. (2012). A graphical tool for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. *Computational Statistics*, *27*, 103–125.
- [53] Lue, H.-H. (2009). Sliced inverse regression for multivariate response regression. *Journal of Statistical Planning and Inference*, *139*, 2656–2664.
- [54] Naik, P. and Tsai, C. L. (2000) Partial least squares estimator for single-index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*, 763–771.
- [55] Nkiet, G.-M. (2008). Consistent estimation of the dimensionality in sliced inverse regression. *Annals of the Institute of Statistical Mathematics*, *60*, 257–271.
- [56] Prendergast, L. A. (2005). Influence functions for sliced inverse regression. *Scandinavian Journal of Statistics*, *32*, 385–404.
- [57] Prendergast, L. A. (2007). Implications of influence function analysis for sliced inverse regression and sliced average variance estimation. *Biometrika*, *94*, 585–601.
- [58] Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in statistics - Theory and methods*, *26*, 2141–2171.
- [59] Saracco, J. (1999). Sliced inverse regression under linear constraints. *Communications in statistics - Theory and methods*, *28*, 2367–2393.

- [60] Saracco, J. (2001). Pooled slicing methods versus slicing methods. *Communications in statistics - Simulation and Computation*, 30, 489–511.
- [61] Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR_α approach. *Journal of Multivariate Analysis*, 96 117–135.
- [62] Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression. *Journal of the American Statistical Association*, 89, 141–148.
- [63] Scrucca, L. (2007). Class prediction and gene selection for DNA microarrays using regularized sliced inverse regression. *Computational Statistics & Data Analysis*, 52, 438–451.
- [64] Setodji, C. M. and Cook, R. D. (2004). K-means inverse regression. *Technometrics*, 46, 421–429.
- [65] Shao, Y., Cook, R.D., Weisberg, S., (2007). Marginal tests with sliced average variance estimation. *Biometrika*, 94, 285-296.
- [66] Shao, Y., Cook, R. D., and Weisberg, S. (2009). Partial central subspace and sliced average variance estimation. *Journal of Statistical Planning and Inference*, 139, 952–961.
- [67] Szretter, M. E. and Yohai, V. J. (2009). The sliced inverse regression algorithm as a maximum likelihood procedure. *Journal of Statistical Planning and Inference*, 139, 3570–3578.
- [68] Weisberg, S., (2002). Dimension reduction regression in R. *Journal of Statistical Software*, Available from <http://www.jstatsoft.org>
- [69] Wen, X., Cook, R.D., (2007). Optimal sufficient dimension reduction in regressions with categorical predictors. *Journal of Statistical Inference and Plannin*, 137, 1961-1979.
- [70] Wu, H.-M. (2008). Kernel sliced inverse regression with applications to classification. *Journal of Computational and Graphical Statistics*, 17, 590–610.
- [71] Ye, Z. and Yang, J. (2010). Sliced inverse moment regression using weighted chi-squared tests for dimension reduction. *Journal of Statistical Planning and Inference*, 140, 3121–3131.

- [72] Ye, Z., Weiss, R.E., (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association*, 98, 968-979.
- [73] Yin, X., Cook, R.D., (2002). Dimension reduction for the conditional k th moment in regression. *Journal of the Royal Statistical Society, Ser. B.* 64, 159-175.
- [74] Yin, X., Cook, R.D., (2003). Estimating central subspaces via inverse third moments. *Biometrika*, 90, 113-125.
- [75] Yin, X. and Bura, E. (2006). Moment-based dimension reduction for multivariate response regression. *Journal of Statistical Planning and Inference*, 136, 3675–3688.
- [76] Zhong, W., Zeng, P., Ma, P., Liu, J.S. and Zhu, Y. (2005). RSIR: Regularized Sliced Inverse Regression for motif discovery. *Bioinformatics*, 21, 4169–4175.
- [77] Zhu, L., Miao, B., and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*, 101, 630–643.
- [78] Zhu, L.-P. and Yu, Z. (2007). On spline approximation of sliced inverse regression. *Science in China Series A: Mathematics*, 50, 1289–1302.
- [79] Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics*, 24, 1053–1068.
- [80] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statistica Sinica*, 5, 727–736.
- [81] Zhu, L. X., Ohtaki, M., and Li, Y. (2007). On hybrid methods of inverse regression-based algorithms. *Computational Statistics*, 51, 2621–2635.