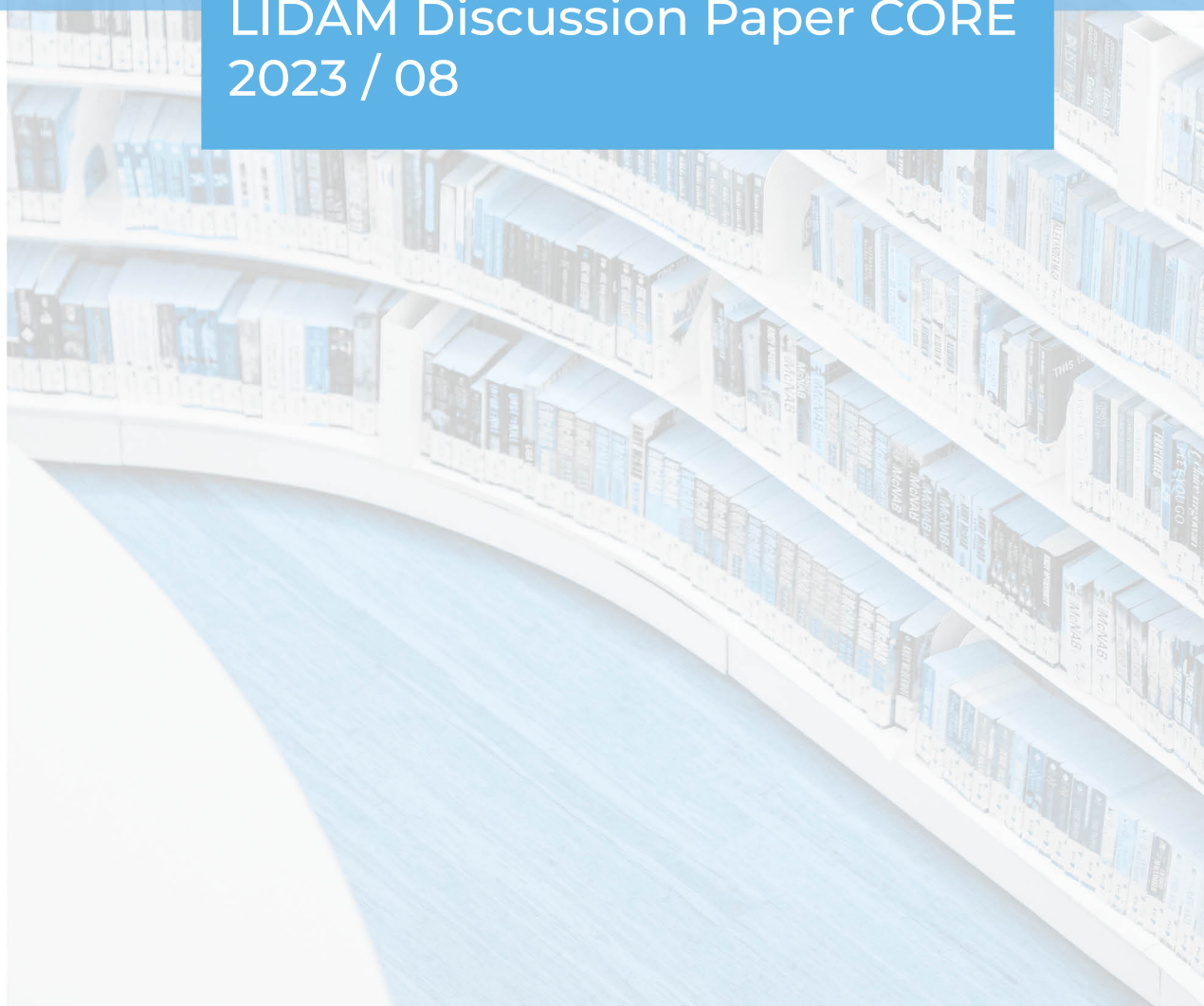
GOODNESS-OF-FIT TEST IN HIGH-DIMENSIONAL LINEAR SPARSE MODELS

Mathieu Sauvenier, Sébastien Van Bellegem

LIDAM Discussion Paper CORE
2023 / 08



UCLouvain

CORE | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

Goodness-of-fit test in high-dimensional linear sparse models

Mathieu Sauvenier Sébastien Van Bellegem

LIDAM/CORE, Université catholique de Louvain, Belgium

March 17, 2023

Abstract

A goodness-of-fit test for the outcome of variable selection in a high dimensional linear model is studied. The test minimizes a moment condition that reflects the sparsity constraint. Testing this constraint is possible thanks to a high dimensional central limit Theorem that is proved under heteroskedasticity. To implement the test a multiple-splitting projection test procedure that has been recently proposed in the literature is employed. Monte Carlo experiments demonstrate the power of the test. A real data application considers the problem of selecting predictors to nowcast quarterly GDP. The empirical results show that it is possible to select a minimal number of variables such that every larger set of variables would pass the goodness-of-fit test.

Keywords: High dimensional model; Sparsity; Goodness-of-Fit; Projection test; Nowcasting

1 Introduction

In high dimensional statistical models, the assumption of sparsity relates to a situation where many covariates are conditionally independent from the dependent variable, given a smaller set of covariates. Finding the appropriate conditioning covariates is the topic of a large literature in statistics and empirical sciences. The purpose of this paper is to study a goodness-of-fit procedure to test the validity of the selected variables given by a model selection procedure.

The literature provides significance test on (a fraction of) high dimensional estimators that takes the form of a penalized least square. This is for instance the case in Bühlmann (2013) for Ridge estimation, in van de Geer et al. (2014) for LASSO, in Meinshausen (2015) for an l_1 penalty, in Gold et al. (2020) for a combination of LASSO and instrumental variables techniques.

The test studied in the current paper does not suppose any particular estimation procedure. Given any estimated selection of significant covariates, the goodness-of-fit test evaluates the null of sparsity. After an identification result that is of an independent interest, the large sample properties of the testing procedure is established under heteroskedasticity. A feasible test is provided and based on a multiple-splitting projection test (Liu et al., 2022). The finite-sample properties of the test are studied through Monte Carlo experiments and an application in macroeconomics is provided, showing the good behavior of the test in finite sample situations.

2 Construction of the test statistics

2.1 Sparsity condition

Let y be a real, univariate dependent variable and x be a set of p real explanatory variables. We assume exogeneity, that is the parameters of the marginal distribution x and the conditional distribution $(y \mid z)$ operate a sequential cut (Engle et al., 1983). The vector of explanatory variables is splitted into two subsets of variables such that $z = (x^{\mathrm{T}}, w^{\mathrm{T}})^{\mathrm{T}}$ where x has dimension k and w has dimension $p - k$. We moreover assume linearity that is there exist parameters β in $\mathbb{R}^k$ and β_w in $\mathbb{R}^{p-k}$ such that the conditional expectation writes

$$E(y \mid x, w) = x^{\mathrm{T}}\beta + w^{\mathrm{T}}\beta_w \, . \tag{1}$$

Sparsity is a modeling situation where only the first term in model (1) is sufficient to define the conditional expectation, that is $E(y \mid x, w)$ does not depend on w. In other words, sparsity is a notion of conditional decorrelation. The following lemma provides a characterization of sparsity that is meaningful for the next testing procedure.

Lemma 1. *Consider the linear model (1) and assume the rank of the matrix $E(xx^{\mathrm{T}})$ is k. Consider the following statements:*

$$w^{\mathrm{T}}\beta_w = 0 \tag{2}$$

$$E\big(wy - wx^{\mathrm{T}}E(xx^{\mathrm{T}})^{-1}E(xy)\big) = 0 \tag{3}$$

$$\beta = E(xx^{\mathrm{T}})^{-1}E(xy) \tag{4}$$

If $\mathrm{range}\big(E(ww^{\mathrm{T}})\big) \cap \mathrm{range}\big(E(wx^{\mathrm{T}})\big) = \{0\}$ *then (2) ⇔ (3) ⇒ (4). If in addition at least a row of* $E(xw^{\mathrm{T}})$ *is nonzero, then (2) ⇔ (3) ⇔ (4).*

The above definition of sparsity is slightly different from the conception that $\beta_w = 0$ in model (1). We rather write the definition in terms of conditional covariance of the explanatory variables. Equivalently we can say that β_w is a nuisance parameter, possibly not identified, and not necessary vanishing.

2.2 Test statistics

From an independent and identically distributed sample of n dependent and explanatory variables, denote by $Y = (y_1, \ldots, y_n)^{\mathrm{T}}$ the vector of observed realizations of y and $Z = (z_1^{\mathrm{T}}, \ldots, z_n^{\mathrm{T}})^{\mathrm{T}}$ is the $n \times p$ matrix of observed explanatory variables. According to the decomposition (1), $Z = (X; W)$ is a partition of Z where $X = (x_1^{\mathrm{T}}, \ldots, x_n^{\mathrm{T}})^{\mathrm{T}}$ is the $n \times k$ submatrix containing the realizations of x whereas $W = (w_1^{\mathrm{T}}, \ldots, w_n^{\mathrm{T}})^{\mathrm{T}}$ contains the realizations of variables w. From Lemma 1 a natural test statistic is given by $T_n := n^{-1} W^{\mathrm{T}}(Y - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y)$ for every n. The next result provides an approximation of the distribution of T_n in high dimension that is a situation where the number of covariates, p, is possibly increasing when the sample size grows.

Proposition 1. *Consider model (1) under sparsity and assume that* $X^{\mathrm{T}}X$ *is invertible almost surely. Let the projection* $P_X = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ *and denote by* $e = (I - P_X)Y$ *the residuals after projection onto* X*. Denote by* $\Omega = E(ee^{\mathrm{T}} \mid X, W)$ *the diagonal conditional covariance matrix of those residuals. For some* $\nu > 2$ *suppose*

$$\max_i E(|e_i|^\nu \mid X, W) \text{ is uniformly bounded} \tag{5}$$

$$\sup_{\phi \in \mathbb{R}^{p-k}} E|\phi W|^\nu < \infty \text{ componentwire} \tag{6}$$

$$\min_i E((W\phi)_i^2 \Omega_{ii}) \geq C > 0. \tag{7}$$

Then the test statistics T_n *is such that, for every* $\phi \in \mathbb{R}^{p-k}$ *satisfying the assumptions,*

$$\sqrt{n}\phi^{\mathrm{T}} T_n \text{ is } AN\Big(0, \frac{1}{n} E(\phi^{\mathrm{T}} W^{\mathrm{T}} \Omega W \phi)\Big). \tag{8}$$

Assumption (5) is a conditional moment restriction on the residuals e_i and Assumption (6) is a moment restriction on the observable variables. If the model is homoskedastic, then $\Omega_{ii} = \sigma^2$ for all i and condition (7) simplifies.

3 A feasible test

3.1 Sparse projection tests

Given the Normal approximation (8) we consider the test statistics given by

$$nT_n^{\mathrm{T}}\phi(\phi^{\mathrm{T}}W^{\mathrm{T}}\Omega_n W\phi)^{-1}\phi^{\mathrm{T}}T_n$$

where $\Omega_n = ee^{\mathrm{T}}/n$ is the empirical counterpart of Ω. Targeting the best power of the test, Huang (2015) finds that the optimal vector ϕ solves the inverse problem $\phi = \{E(W^{\mathrm{T}}\Omega W)\}^{-1}E(T_n|\mathrm{H}_1)$, where H_1 denotes the alternative that the sparse subset of covariates is not included in X. Given that under the null our model is sparse, we circumvent the inverse problem and select $\phi \in \mathbb{R}^{(p-k)}$ that solves that minimizes the following penalized quadratic form: following optimisation:

$$\phi^{\mathrm{T}}W^{\mathrm{T}}\Omega_n W\phi + g_\lambda(\phi) = \frac{1}{n}\|\phi^{\mathrm{T}}W^{\mathrm{T}}e\|^2 + g_\lambda(\phi) \tag{9}$$

where g_λ is a penalty function.

3.2 Multiple data splitting

In order to easily find the distribution we conduct the inference on ϕ and T_n on separate samples. We use $r < n$ data to get an estimate of ϕ from (9) that we denote by ϕ_r. We then use the $s = n-r$ complementary data to compute $T_{si} = w_i(\delta_i - x_i^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^T)Y/s$ where δ_i is a vector of zeros, except the ith component where it is one. Proposition 1 shows that $\phi_r^{\mathrm{T}}T_{si}$ is approximately Normal under the null conditionally on ϕ_r. Therefore we consider the test statistics $U_{rs} = \sqrt{s}\phi_r^{\mathrm{T}}T_s/\hat{\gamma}$, where $\hat{\gamma}^2$ is the sample variance of $\{\phi_r^{\mathrm{T}}T_{si}; 1 \leq i \leq s\}$. If ζ denotes a standard Normal variable, the $p-$value of the bilateral test base on U_{rs} is therefore approximately given by $2(1 - P(\zeta \leq |U_{rs}|))$ for s sufficiently large and r fixed.

Splitting the data reduces the power of the test since it reduces the sample size. In order to increase the power of the test and to stay robust with respect to a particular split, Liu et al. (2022) operate multiple splits of the data. They consider m permutations of $\{1, ..., n\}$, leading to m partitions of the sample. On each partition of $r + s$ data we compute the test statistics. We finally get m p-values of the test that we denote by $\pi_1, \ldots, \pi_m$.

The $p-$values $\pi_1, ..., \pi_m$ may be dependent. Liu et al. (2022) prove that there are *exchangeable* that is $(\pi_1, ..., \pi_m)$ is identically distributed to $(\pi_{\rho(1)}, ..., \pi_{\rho(m)})$ for any permutation ρ of the indices $\{1, ..., m\}$. Exchangeability holds true under both the null and the alternative.

To combine the $p-$values we define ξ_j such that $P(\zeta \leq \xi_j) = \pi_j$ for all $j = 1, \ldots, m$. Under the null, $\xi_1, \ldots, \xi_m$ are marginally standard Normal and jointly exchangeable with a positive correlation between any disjoint pair. If in addition we *assume* that $\xi_1, ..., \xi_m$ are multivariate Normal, the setup is identical to the setup of Follmann and Proschan (2012) who provides a test of location that exactly control for the type I error.

Following Follmann and Proschan (2012) we assume for all $j = 1, ..., m$ that $\xi_j = V + \epsilon_j$ where $V \sim \mathcal{N}(0, \sigma_V^2)$, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$, that $(V, \epsilon_1, ...\epsilon_m)$ are independent and that $\sigma_V^2 + \sigma_\sigma^2 = 1$. Under this last set of assumptions, we have that the correlation between any disjoint pairs of ξ_js is $\kappa = 1 - \sigma_\epsilon^2 \geq 0$ (Follmann and Proschan, 2012). Moreover those assumptions are useful to control the estimation of κ and to derive the exact distribution of $\xi_{\hat{\kappa}} = \sqrt{m/(1 + (m-1)\hat{\kappa})}\overline{\xi}$, where $\overline{\xi}$ is the average of ξ_js and $\hat{\kappa}$ is an estimator of the correlation κ. In particular Follmann and Proschan (2012) proves that for the conservative estimate of κ that is $\hat{\kappa}(\tau) = \max(0, (m-1)s_\xi^2/a)$, where s_ξ^2 is the empirical variance of $\{\xi_1, ..., \xi_m\}$ and a solves $P(x > a) = \tau$ for x being a χ_{m-1}^2 random variable, we have under the null that

$$P(\xi_{\hat{\kappa}(\tau)} > c) = \int_0^{\frac{a}{1-\kappa}} P\left(\zeta^2 > c^2 \frac{1 + (m-1)(1 - (1-\kappa)t/a)}{1 - (m-1)\kappa}\right) h_{m-1}(t)dt \\ + P\big(\zeta^2 > c^2/(1 + (m-1)\kappa)\big)\bar{H}_{m-1}\big(\frac{a}{1-\kappa}\big) \quad (10)$$

where $\bar{H}_{m-1}$ is $1 - P(x \leq c)$ and h_{m-1} is the density of χ_{m-1}^2.

Since the probability in Equation (10) depends on the unknown κ, we may protect against the least favourable κ by considering the bound $\sup_{\kappa \in [0,1)} P(\xi_{\hat{\kappa}(\tau)} > c)$ (Follmann and Proschan, 2012).

This ensures that the type I error is controlled for *any* κ. In practice solving this maximisation problem can be achieved by solving Equation (10) via numeric integration for various κ .

To control the type I error in practice we fix m and choose $c = c^\star$ and find τ such that $\sup_{\rho\in[0,1)} P(\xi_{\hat{\kappa}(\tau)} > c^\star) = \alpha$ where α is the probability of type I error. Table 1 gives values for τ when $c^\star = 1.96$ and $\alpha = 0.05, 0.1$ for a two-sided test.

m	2	5	10	20	40	100	1000	10000
$\tau_{0.05}$	.25	.25	.20	.20	.15	.15	.10	.05
$\tau_{0.1}$	.74	.535	.45	.39	.345	.295	.215	.17

TABLE 1: *Statistical table. Value of τ that, for the given m, ensures the probability of type I error is bounded by $\alpha = 0.05, 0.1$ for a two-sided test with critical value $c^\star = 1.96$.*

4 Finite sample properties

The final sample properties of the test are studies through Monte Carlo experiments. Consider $n = 100$ independent and identically data and $p = 200$ covariates generated such that $Z \sim \mathcal{N}_p(0, \Sigma)$ and $Y|Z \sim \mathcal{N}_n(Z\gamma, \gamma^{\mathrm{T}}\Sigma\gamma/S)$ where S is the signal-to-noise ratio and γ is generated such that there are 10 randomly selected elements of γ that are assigned a value of 1 while the other elements of γ are assigned the value zero. Next, for $k = 1, ..., 20$ we construct $X \in \mathbb{R}^{n\times k}$ and $W \in \mathbb{R}^{n\times(p-k)}$ using the following settings.

If $k \leq 10$, X contains the k first columns of Z that are associated with a non-null element of γ, and W contains the complementary columns of Z. If $k \geq 10$, X contains the 10 columns of X associated with a non-null element of β and $k-10$ other columns that are randomly selected. Again W contains the complementary columns of Z. Below, for any given k, we refer to this selection of covariates as the *Oracle selection*.

We now test the null $H_0 : W^\top\beta_w = 0$ following the feasible testing procedure described in the previous section, with calibration values $m = 100$, $\alpha = 0.1$ and $c^\star = 1.96$. We run the experiment in three different scenarios that are

1. The covariance of Z is $\Sigma_{i,j} = 0.5^{|i-j|}$ and the signal-to-noise ratio is $S = 6$;

2. The covariance of Z is $\Sigma_{i,j} = 0.5^{|i-j|}$ and the signal-to-noise ratio is $S = 2$;

3. The covariance of Z is $\Sigma = U^{\mathrm{T}}\Lambda U$ where $\Lambda = \mathrm{diag}(10/\sqrt{j})$, U contains the eigenvectors of the matrix $(0.5)^{|i-j|}$ and the signal-to-noise ratio is $S = 2$.

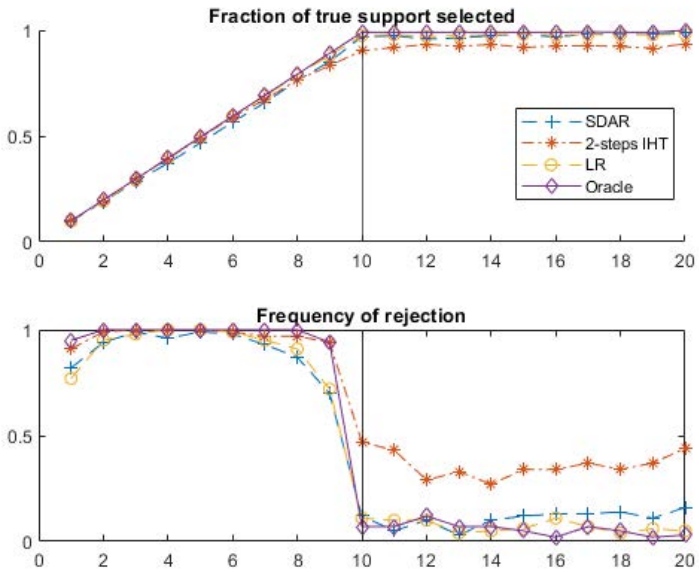


FIGURE 1: *Scenario 1. (Top) Fraction of the true support of β selected by the three estimators as well as by the Oracle estimator. (Bottom) Frequency of rejection of H_0 based on the selection operated by four estimators.*

The results are displayed in Figures 1, 2 and 3 for three selection methods that are explained below.

We now describe the three selection methods used in Figures 1, 2 and 3. The method SDAR stands for 'Support Detection And Root' (Huang et al., 2018). It approximates an l_0-penalized least squares. The two-stages IHT is designed to approximate l_0 constrained least squares programs (Jain et al., 2014). Finally the Loaded-Rifle (LR) is a directional l_0 constrained estimator studied in Sauvenier and Van Bellegem (2023). The results given by the Oracle selection are also displayed on the pictures.

The results call for the following comments. First the test is powerful in all considered scenarios. Second, decreasing the signal-to-noise ratio inflates type II error when the selection contains most of the relevant variables. Third, when the covariance matrix becomes more complicated to estimate (Scenario 3) type II error also inflates when the selection contains most of the important variables. In both cases these losses of power may be attributed to the increased difficulty of estimating the

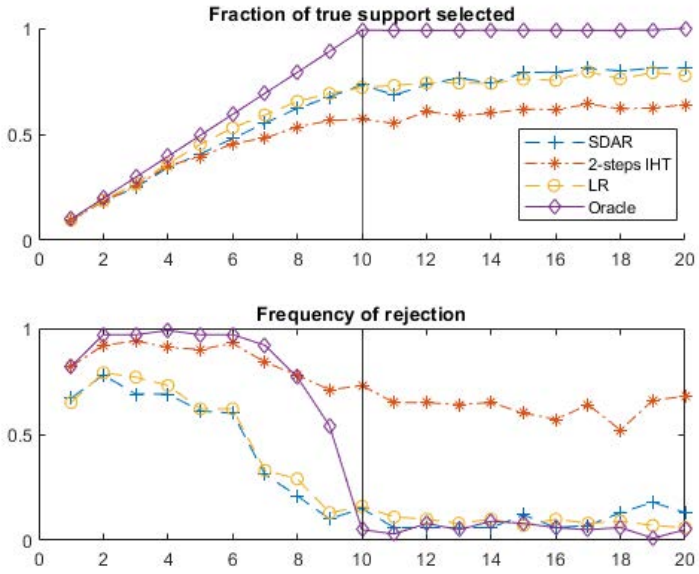


FIGURE 2: *Scenario 2. (Top) Fraction of the true support of β selected by the three estimators as well as by the Oracle estimator. (Bottom) Frequency of rejection of H_0 based on the selection operated by four estimators.*

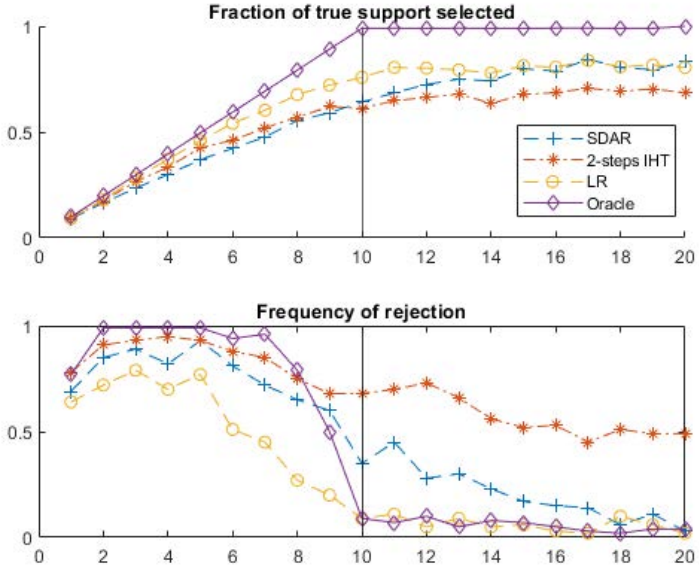


FIGURE 3: *Scenario 3. (Top) Fraction of the true support of β selected by the three estimators as well as by the Oracle estimator. (Bottom) Frequency of rejection of H_0 based on the selection operated by four estimators.*

optimal direction in (9). Four, the test procedure adequately controls type I error. Finally, we also note the presence of non-marginal type II errors when $k = 1, 2$.

5 Application to nowcasting

Key economic variables such as GDP, aggregate private consumption or sectorial added value are published by economic institutions at low frequency (quarterly) and with some delay. Early predictions and estimates may be achieved by using information provided in data published at higher frequencies (monthly). This problem is known as *nowcasting* and is common and important in many institutions such as central banks and research centres (Giannone et al., 2008). One line of research in nowcasting focuses on the use of large and high-dimensional datasets. However it is argued that an increased number of predictors may hurt the prediction (Caggiano et al., 2011). Selecting a subset of variables from a large dataset may lead to better forecast, but there is no consensus on how to select the predictor. Penalized least squares programs are definitely one of the main methods used by practitioners (Cepni et al., 2020).

We consider GDP nowcasting in Belgium. Four procedures for variable selection are used to select sets of predictors of the first difference of the seasonally adjusted log GDP measured in constant price. The potential predictors are 471 monthly time series that are used by economic forecasting services (Attanasi et al., 2020). They contain a sample of business and consumer surveys, economic indices as well as real, financial and monetary variables. The sample period range from the first quarter of 2004 to the second quarter of 2022 both for monthly and quarterly data. All the data are downloaded from Macrobond's data services.

When monthly time series are available quarterly time series are constructed by taking the average over each quarter. We transform the obtained time series as in Piette et al. (2014), that is the logarithm is taken for every strictly positive times series and apply a first difference to every times series displaying a unit-root. Finally, we z-score all the covariates.

We run Elastic-net (Zou and Hastie, 2005), SDAR, 2-steps IHT and LR to select at least k predictors ($1 \leq k \leq 48$) and compute the $p-$values for all the selections with $m = 1000$ and probability of type I error bounded by 0.05. We also compute the R^2 of the post selection linear regression for each selection and plot the $p-$value against the R^2 in Figure 4.

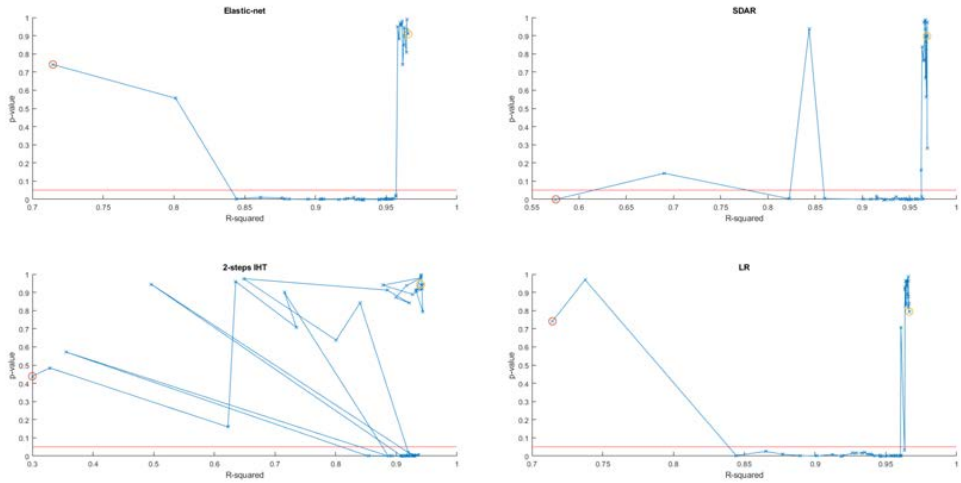


FIGURE 4: *Results of the goodness-of-fit tests. $p-$value plotted against R^2 for* 48 *selections operated by the estimator indicated on the graph. The model with one variable is surrounded by a red circle, the model with 48 variables is surrounded by an orange circle. Each dot is connected by a line to the dots corresponding to selections one $k-1$ and $k+1$ selected variables. The red horizontal line corresponds to a $p-$value of* 0.05.

Figure 4 interestingly shows there are $\bar{k} \in \mathbb{N}$ such that for all $k \geq \bar{k}$ the selections have high $p-$value and high R^2. This is a desirable feature that is expected from a selection encompassing all the important variables of a linear model. Moreover for each estimator but IHT, the selections for $5 < k < \bar{k}$ are consistently rejected by the test. For $k \geq \bar{k}$ they are consistently no longer rejected. This behavior is desirable since it displays the consistency of the test across the sequence of selection and indicates the test is indeed able to capture a qualitative change in the selections. It also suggests considering as fake negatives the non-rejection of H_0 for selections operated by Elastic-net and LR for $k = 1, 2$. The increased probability of fake negative for low k had already been observed in controlled experiments in the previous section. Finally we observe there is coherency between the results of the four different estimators since for each estimator $\bar{k} \in \{31, ..., 35\}$. Overall, the feasible test provides coherent and meaningful information concerning selection of variables on this real dataset. It suggests choosing the set of predictors of the Belgian GDP that correspond to selections with $k \geq \bar{k}$. This leads to a non-trivial subset since operating the selection via a cross-validated Elastic-net only leads to selecting 11 covariates, a model that is rejected by our test and is far bellow $\bar{k}$.

6 Conclusion

The contribution of the paper are twofold. First an identification result for the parameter of a sparse linear model is proved. Second, a test based on multiple splitting is studied, both for large sample sizes and for small samples. An original application to the nowcasting of Belgian GDP shows non standard results in the selection of relevant economic variables.

A Proofs

A.1 Proof of Lemma 1

All variables are vectors in the Hilbert space $L^2(\Omega, \mathcal{A}, \mathbb{P})$. The conditional expectation in Equation (1) is the orthogonal projection of y onto the hyperplane spanned by x and w. Denote this orthogonal projection by P_{xw}, and Q_{xw} is the projection onto the subspace that is orthogonal to $(x^{\mathrm{T}}, w^{\mathrm{T}})^{\mathrm{T}}$. Therefore we have the orthogonal decomposition $y = P_{xw}y + Q_{xw}y$. Using the parametrization of $P_{xw}y$ in Equation (1) and the projection Theorem (Halmos, 1951), we get the Normal equations

$$E(x(y - x^{\mathrm{T}}\beta + w^{\mathrm{T}}\beta_w)) = 0 \tag{11}$$

and

$$E(w(y - x^{\mathrm{T}}\beta + w^{\mathrm{T}}\beta_w)) = 0\ . \tag{12}$$

To prove (2) $\Rightarrow$ (3) we take $E(w^{\mathrm{T}}\beta_w) = 0$ in both equations and substitute β from (11) in (12). To prove (3) $\Rightarrow$ (2), isolate β in (11) and substitute in (12) to get, from (3), $0 = (E(ww^{\mathrm{T}}) - E(wx^{\mathrm{T}})E(xx^{\mathrm{T}})^{-1}E(xw^{\mathrm{T}}))\beta_w$ .. From the range condition stated in the Lemma we conclude that $E(ww^{\mathrm{T}})\beta_w = 0$. If we premultiply the last by β_w^{T} we get $E(w^{\mathrm{T}}\beta)^2 = 0$ and therefore (2) almost surely. From (11) we get directly (2) $\Rightarrow$ (4). Conversly from (11) and (4) we get $E(xw^{\mathrm{T}})\beta_w = 0$. Given that at least a line in $E(xw^{\mathrm{T}})$ is non zero, it is necessary that $w^{\mathrm{T}}\beta_w = 0$.

A.2 Proof of Proposition 1

Let $P_{XW} = E(Y \mid X, W)$ denote the orthogonal projection onto (X, W). Under the null the conditional expectation does not depend on W (therefore $P_{XW} = P_X$) and it is linear in its

parameters. If we write $e = Y - P_X Y$ then $T_n = W^{\mathrm{T}} e/n$ is obviously orthogonal to span(X, W). It is therefore conditionally zero mean with conditional variance given in (8). The Normal approximation of $\phi^{\mathrm{T}} T_n$ follows from the Lindeberg-Feller central limit theorem (e.g. Theorem and Corollary 1.9.3 in Serfling (1980)) for any ϕ in $\mathbb{R}^{p-k}$ under the stated assumptions.

We provide the details here. Observe

$$\phi^{\mathrm{T}} T_n = \frac{1}{n} \sum_1^n \phi W_i e_i \ .$$

Then because $W_i \in \mathbb{R}^{p-k}$ and $p = p(n)$, we write $U_{i;n} = \phi W_i e_i$.

Lindeberg-Feller Theorem states that, if $\sum_i E|U_{i;n}|^\nu = o(B_n^\nu)$ as $n \longrightarrow \infty$ for some $\nu > 2$, then $\sum_i U_{i;n}$ is $AN(0, B_n^2)$. In our setting $B_n^2 = E(\phi^{\mathrm{T}} W^{\mathrm{T}} \Omega W \phi)$ where $\Omega = E(ee^{\mathrm{T}} \mid X, W)$. Thus

$$\begin{aligned} \sum_i E(|U_{i;n}|^\nu \mid X, W) &= \sum_i E(|\phi W_i e_i|^\nu \mid X, W) \\ &= \sum_i |\phi W_i|^\nu E(|e_i|^\nu \mid X, W) \\ &< \sum_i |\phi W_i|^\nu \end{aligned}$$

by Assumption (5) and

$$\frac{1}{n} \sum_i E|U_{i;n}|^\nu < \frac{1}{n} \sum_i E|\phi W_i|^\nu$$

that is uniformly bounded under Assumption (6). It is now sufficient to show that B_n^2/n diverges since $\nu > 2$. From the independence assumption implies that Ω is diagonal. Therefore

$$\begin{aligned} B_n^2 &= \sum_i \sum_j E((W\phi)_i (W\phi)_j \Omega_{ij}) \\ &= \sum_i \sum_j E((W\phi)_i^2 \Omega_{ii}) \\ &\geq Cn^2 \end{aligned}$$

by Assumption (7).

References

G. Attanasi, V. Bodart, F. Courtoy, S. Fontenay, N. Lachapelle, A. Ounnas, M. Sauvenier, et al. Perspectives économiques 2020-2021. Technical report, UCLouvain, LIDAM/IRES, 2020.

P. Bühlmann. Statistical significance in high-dimensional linear models. *Bernoulli*, 19:1212–1242, 2013.

G. Caggiano, G. Kapetanios, and V. Labhard. Are more data always better for factor analysis? results for the euro area, the six largest euro area countries and the uk. *J Forecast*, 30:736–752, 2011.

O. Cepni, I. E. Guney, and N. R. Swanson. Forecasting and nowcasting emerging market gdp growth rates: The role of latent global economic policy uncertainty and macroeconomic data surprise factors. *J Forecast*, 39:18–36, 2020.

R. Engle, D. Hendry, and J.-F. Richard. Exogeneity. *Econometrica*, 51:277–304, 1983.

D. Follmann and M. Proschan. A test of location for exchangeable multivariate normal data with unknown correlation. *J Multivar Anal*, 104:115–125, 2012.

D. Giannone, L. Reichlin, and D. Small. Nowcasting: The real-time informational content of macroeconomic data. *J Monet Econ*, 55:665–676, 2008.

D. Gold, J. Lederer, and J. Tao. Inference for high-dimensional instrumental variables regression. *J Econom*, 217:79–111, 2020.

P. Halmos. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. Chelsea, New York, 1951.

J. Huang, Y. Jiao, Y. Liu, and X. Lu. A constructive approach to L_0 penalized regression. *J Mach Learn Res*, 19:1–37, 2018.

Y. Huang. *Projection test for high-dimensional mean vectors with optimal direction*. PhD thesis, 2015. Thesis, Department of Statistics, The Pennsylvania State University.

P. Jain, A. Tewari, and P. Kar. On iterative hard thresholding methods for high-dimensional m-estimation. In Z. Ghahramani and et al., editors, *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014.

W. Liu, X. Yu, and R. Li. Multiple-splitting projection test for high-dimensional mean vectors. *J Mach Learn Res*, 23(71):1–27, 2022.

N. Meinshausen. Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *J R Stat Soc Series B*, 77:923–945, 2015.

C. Piette, G. Langenus, et al. Using brel to nowcast the belgian business cycle: the role of survey data. *Econ Rev*, pages 75–98, 2014.

M. Sauvenier and S. Van Bellegem. Direction identification and minimax estimation by generalized eigenvalue problem in high dimensional sparse regression. (LIDAM CORE 2023/05), 2023.

R. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann Statist*, 42:1166–1202, 2014.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J R Stat Soc Series B*, 67:301–320, 2005.