This is the accepted version of a paper presented at 31st European Signal Processing Conference, EUSIPCO 2023, Helsinki, Finland.

Citation for the original published paper:

Borpatra Gohain, P., Jansson, M. (2023)
Model Selection in High-Dimensional Block-Sparse General Linear Regression
In: (pp. 1928-1932-).
https://doi.org/10.23919/EUSIPCO58844.2023.10289738

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-339924
Model Selection in High-Dimensional Block-Sparse General Linear Regression

Prakash B. Gohain*, Magnus Jansson†
*†Division of Information Science and Engineering, KTH Royal Institute of Technology
*Ericsson
Stockholm, Sweden
prakash.b.gohain@gmail.com, janssonm@kth.se

Abstract—Model selection is an indispensable part of data analysis dealing very frequently with fitting and prediction purposes. In this paper, we tackle the problem of model selection in a general linear regression where the parameter matrix possesses a block-sparse structure, i.e., the non-zero entries occur in clusters or blocks and the number of such non-zero blocks is very small compared to the parameter dimension. Furthermore, a high-dimensional setting is considered where the parameter dimension is quite large compared to the number of available measurements. To perform model selection in this setting, we present an information criterion that is a generalization of the Extended Bayesian Information Criterion-Robust (EBIC-R) and it takes into account both the block structure and the high-dimensionality scenario. We name it Generalized EBIC-R (GEBIC-R). The analytical steps for deriving the GEBIC-R are provided. Simulation results show that the proposed method performs considerably better than the existing state-of-the-art methods and achieves empirical consistency at large sample sizes and/or at high-SNR.

Index Terms—Model selection, block-sparsity, compressed sensing, information criterion, orthogonal matching pursuit.

I. INTRODUCTION

Selecting the best model/subset in the high-dimensional (HD) linear regression has been an active research topic for a long time now. In this context, methods based on Information Criterion (IC) have played a pivotal role ever since Akaike proposed the famous Akaike IC [1]. In the present era, popular IC-based methods for model selection in the HD setting include extended Bayesian IC (EBIC) [2], extended Fisher IC (EFIC) [3], and extended BIC-Robust (EBIC-R) [4], [5]. Apart from the IC-based methods, there are other non-IC-based model selection approaches in the HD regime. They include methods based on hypothesis testing framework such as the Residual-Ratio-Thresholding (RRT) [6] and the Multi-Beta-Test (MBT) [7]. Some recent methods also include the significance test in non-monotonic predictor sequences generated by LASSO [8]. Another popular method is cross-validation (CV) [10], [11]. However, CV-based procedures can be computationally intensive, and their performance in the HD setting is not satisfactory [12]–[14].

In this paper, we specifically consider model selection in a general linear regression where the nonzero coefficients in the parameter matrix occur in clusters (or groups). Such signals are referred to as block-sparse [15]–[17]. Block-sparsity inherently arises in a variety of scenarios. For example in multi-band signals [18], [19], in the recovery of signals from compressed microarray measurements [20], and in the multiple measurement vector (MMV) problem [21]–[24]. Furthermore, as shown in [15] and [16], the block-sparsity model can be used to handle the issue of sampling signals that lie in a union of subspaces [25], [26].

A recent method for model selection in block-sparse HD linear regression is the Generalized RRT (GRRT) [27]. GRRT is an extension of RRT [6] developed to treat the block-sparse structure in linear regression. The authors also present a new approach that allows GRRT to perform model selection in non-monotonic predictor sequences generated by LASSO [28]. However, to the best of our knowledge, there are no existing IC-based methods designed to take into account the block structure during model selection. Hence, in their current form, they cannot be applied directly without tailoring them to incorporate the block nature of the underlying linear model into the criterion. In this paper, the main goal is to develop an IC-based model selection method for the general linear regression model assuming a block structure and HD setting.

Notations: matrices and vectors are denoted by boldface letters. The notation $(\cdot)^T$ stands for transpose. $I_N$ is an $N \times N$ identity matrix. $\Pi(A) = A(A^T A)^{-1}A^T$ represents the orthogonal projection matrix on the column space of $A$ and $\Pi^\perp(A) = \Pi_N - \Pi(A)$ the orthogonal projection matrix on the null space of $A^T$. The notation $|X|$ denotes the determinant of the matrix $X$, $\|\cdot\|_2$ denotes the Euclidean norm and $\|\cdot\|_F$ the Frobenius norm. $X \sim \mathcal{N}(\mu, C)$ signifies a Gaussian distributed random variable with mean $\mu$ and covariance matrix $C$. The symbol $\otimes$ represents the Kronecker product and vec$(A)$ the vectorization of the matrix $A$. Further, $\text{card}(S)$ denotes the cardinality of the set $S$ and $O(\cdot)$ denotes the standard Big-O notation.

II. PROBLEM STATEMENT

Technically there can be four different linear regression structures depending on the configuration of the parameter matrix (or vector). They are (a) single measurement vector (SMV), (b) block single measurement vector (BSMV), (c)
multiple measurement vector (MMV) and (d) block multiple measurement vector (BMMV). For example, as mentioned in [27], SMV models are used in wireless signal detection [29], MMV models in Electroencephalogram (EEG) [30], BSMV models in multi pitch estimation [31] and BMMV models in face recognition [32]. Here, we consider the BMMV model, since it is the general setting and the rest of the models are special cases of BMMV. The BMMV model is as follows:

\[ Y = AX + W, \]  

where \( Y \in \mathbb{R}^{N \times L} \) is the observed response matrix, \( A \in \mathbb{R}^{N \times p} \) is the design matrix, \( X \in \mathbb{R}^{p \times L} \) is the unknown parameter matrix and \( W \in \mathbb{R}^{N \times L} \) is the noise/error matrix, whose elements are assumed to be i.i.d. and \( W[i, j] \sim N(0, \sigma^2) \). The \( p \) rows of \( X \) are divided into \( p_B = p / L_B \) unique blocks of equal size \( L_B \). Each of these \( p_B \) blocks of size \( L_B \times L \) is non-zero or zero at once. The block size \( L_B \) is assumed to be known a-priori. The \( j \)th block consists of the rows of \( X \) indexed by \( I_B = \{ (j - 1)L_B + 1, (j - 1)L_B + 2, \ldots, jL_B \} \). We denote the true block support of \( X \) as \( S_B = \{ j : X[I_B, :] \neq 0_{k_B L_B} \} \) where \( j \in \{ 1, 2, \ldots, p_B \} \). Also, \( X \) is assumed to be block-sparse such that \( K_B = \text{card}(S_B) \ll p_B \).

Table I shows different linear regression structures. The goal of model selection herein is estimating \( S_B \) given \( Y \) and \( A \).

The model selection procedure can be divided into two stages: (i) In the first stage, we pick a competent set of candidate models using an appropriate predictor/subset selection algorithm up to maximum cardinality \( K \) under the assumption that \( K_B \leq K \ll N \). (ii) In the second stage, we estimate the true model using a suitable model selection criterion. Let us denote \( I_B \) as the block support of a candidate model such that \( \text{card}(I_B) = k_B \), where \( k_B \in \{ 1, 2, \ldots, p_B \} \). Then we can reformulate the linear model in (1) as

\[ \mathcal{H}_{I_B} : Y = A_{I_B} X_{I_B} + W_{I_B}, \]

where \( \mathcal{H}_{I_B} \) signifies the hypothesis that the data \( Y \) is actually produces in accordance with (2), \( A_{I_B} \in \mathbb{R}^{N \times (k_B L_B)} \) is the sub-matrix consisting of columns from the known matrix \( A \) with block support \( I_B \subseteq \{ 1, 2, \ldots, p_B \} \), \( X_{I_B} \in \mathbb{R}^{k_B L_B \times L} \) is the corresponding unknown parameter coefficient matrix, and \( W_{I_B} \in \mathbb{R}^{N \times L} \) is the associated noise matrix.

### III. PROPOSED METHOD

In this section, we provide the necessary steps to derive the GEBIC\(_R\). Note that while EBIC\(_R\) [4], [33] is for model selection in SMV scenarios in the absence of any block structure, GEBIC\(_R\) is a generalization (or extension) of EBIC\(_R\) to perform model selection in block general linear regression (e.g. BMMV scenarios which is the most general setting). The generalization involves some crucial steps that are not trivial and affect the performance drastically. The analysis assumes the following property of the design matrix \( A \) [34]–[36]

\[
\lim_{N \to \infty} \{ N^{-1}(A_{I_B}^T A_{I_B}) \} = M_{I_B},
\]

where \( M_{I_B} \) is a \((k_B L_B \times k_B L_B)\) positive definite matrix and bounded as \( N \to \infty \). The assumption in (3) holds true in many cases but not all (see [34], [37] for more details).

To reformulate the GEBIC\(_R\) for the BMMV model, we first reformulate the linear model in (2) into vector form as

\[ \text{vec}(Y) = I_L \otimes A_{I_B} \text{vec}(X_{I_B}) + \text{vec}(W_{I_B}). \]

This step allows us to utilize the same derivation steps as in EBIC\(_R\) [4] without the need to carry out the analysis from scratch. Also, (4) is technically equivalent to (2), hence we do not alter the underlying original linear model but just restructure it for our convenience. Now, let \( y = \text{vec}(Y) \in \mathbb{R}^{NL \times 1}, \)

\[ \bar{A}_I = I_L \otimes A_{I_B} \in \mathbb{R}^{NL(k_B L_B) \times L}, \]

\[ x_I = \text{vec}(X_{I_B}) \in \mathbb{R}^{i \times NL}, \]

\[ \varepsilon_I = \text{vec}(W_{I_B}) \in \mathbb{R}^{i \times NL}. \]

The elements of \( \varepsilon_I \) are i.i.d. and \( \varepsilon_I \sim N(0, \sigma_I^2 I_{NL}) \). Then, we can rewrite (4) as

\[ \mathcal{H}_I : y = \bar{A}_I x_I + \varepsilon_I, \]

where \( I \subseteq \{ 1, 2, \ldots, pL \} \). Then the pdf of \( y \) under \( \mathcal{H}_I \) is

\[ p(y | \theta_I, \mathcal{H}_I) = \frac{\exp(-\frac{1}{2} \| y - \bar{A}_I x_I \|^2_2 / 2\sigma_I^2)}{(2\pi\sigma_I^2)^{NL/2}}, \]

where \( \theta_I = [x_I^T, \sigma_I^{-2}]^T \) comprises of all the unknown parameters of the model under \( \mathcal{H}_I \). The maximum likelihood estimates (MLE) \( \hat{\theta}_I = [x_I^T, \sigma_I^{-2}]^T \) are obtained as [38]

\[ \hat{x}_I = \left( \bar{A}_I^T \bar{A}_I \right)^{-1} \bar{A}_I^T y \quad \hat{\sigma}_I^2 = \frac{y^T \Pi_I^\perp (\bar{A}_I) y}{NL}, \]

GEBIC\(_R\) is derived under the Bayesian framework of model selection. We follow similar steps as in EBIC\(_R\) [4], [33], but incorporate the block structure into it. Let us denote the prior pdf of the parameter vector \( \theta_I \) as \( p(\theta_I | \mathcal{H}_I) \), the marginal of \( y \) as \( p(y | \mathcal{H}_I) \) and the prior probability of the model with support \( I \) as \( \Pr(\mathcal{H}_I) \). Then the MAP estimate of the true support \( I \subseteq \{ 1, 2, \ldots, pL \} \) is equivalently given by [34], [36]

\[ \hat{S}_{MAP} = \arg \max \left\{ \ln p(y | \mathcal{H}_I) + \ln \Pr(\mathcal{H}_I) \right\}. \]

Applying a second order Taylor series expansion approximation of \( \ln p(y | \mathcal{H}_I) \) is obtained under the presumption that \( N \) is large or/and SNR is high (see [34], [36] for details)

\[ \ln p(y | \mathcal{H}_I) \approx \ln p(y | \hat{\theta}_I, \mathcal{H}_I) + \ln p(\hat{\theta}_I | \mathcal{H}_I) + \frac{k_B L_B + 1}{2} \ln(2\pi) - \frac{1}{2} \ln \| \hat{F}_I \|. \]

Here, \( \hat{F}_I \) is the sample Fisher information matrix [38] under \( \mathcal{H}_I \) evaluated at the MLE which gives us (see [34], [36])

\[ \hat{F}_I = \begin{bmatrix} \frac{1}{\sigma_I^2} \hat{A}_I^T \hat{A}_I & 0 \\ 0 & \frac{N}{\sigma_I^2} \end{bmatrix}. \]
From the linear model in 5 we have

\[ -2 \ln p(y|\hat{\theta}_I, \mathcal{H}_I) = N L \ln \hat{\sigma}_B^2 + \text{const.} \]  \hspace{1cm} (11) \]

Now, using (11), it is possible to rewrite (9) as

\[ -2 \ln p(y|\mathcal{H}_I) \approx NL \ln \hat{\sigma}_B^2 + \ln |\hat{F}_I| - 2 \ln p(\hat{\theta}_I|\mathcal{H}_I) 
- k_B L_B L \ln 2 \pi \] \hspace{1cm} (12) \]

Furthermore, the prior term in (9), i.e., \( \ln p(\hat{\theta}_I|\mathcal{H}_I) \), is ignored under the pretext that it is flat and uninformative. Thus, discarding the constants and the terms not dependent on the block model dimension \( k_B \), we can equivalently reformulate the MAP-based model estimate as

\[ \hat{S}_{\text{MAP}} = \arg \min_{\mathcal{I}} \left\{ NL \ln \hat{\sigma}_B^2 + \ln |\hat{F}_I| - k_B L_B L \ln 2 \pi \right\} \]. \hspace{1cm} (13) \]

GEBIC_{R} is derived from (13) with some further modifications and approximations. The two key terms that require further analysis are \( \ln |\hat{F}_I| \) and the prior term \( \ln \|Q\| \). First, we perform normalization of \( \hat{F}_I \) under both large-\( N \) and high-SNR assumption. For this we factorize the \( \ln |\hat{F}_I| \) term in a similar manner as performed in [4], [33], [36].

\[ \ln |\hat{F}_I| = \ln \left| \left[ \frac{N}{L_B} \right] \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right)^{k_B L_B L} \right| \]
\[ = \ln \left| Q \right| + \ln \left[ \frac{N}{L_B} \right] \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right)^{k_B L_B L} \]. \hspace{1cm} (14) \]

The objective here is to choose a suitable \( Q \) matrix that normalizes \( \hat{F}_I \) such that the second term in (14) is \( O(1) \), i.e., it should be bounded as \( N \to \infty \) and/or \( \sigma^2 \to 0 \). To achieve this purpose, we choose the following \( Q^{-1/2} \) matrix [4]

\[ Q^{-1/2} = \begin{bmatrix} \frac{L_B}{N} & \sqrt{\frac{L_B}{\hat{\sigma}_B^2} k_B L_B L} & 0 \\ 0 & \sqrt{\frac{L_B}{N} \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right) k_B L_B L} & 0 \\ 0 & 0 & \sqrt{\frac{L_B}{N} \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right) k_B L_B L} \end{bmatrix} \] \hspace{1cm} (15) \]

where \( \hat{\sigma}_0^2 = \|y\|_2^2/NL \). Also for the considered generating model (5), \( \hat{\sigma}_0^2 \to \text{const.} \) as \( N \to \infty \) and/or \( \sigma^2 \to 0 \). Two important points to note here regarding the choice of the \( Q^{-1/2} \) matrix are: (i) The ratio \( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \) is introduced to normalize the \( \hat{F}_I \) w.r.t. \( \sigma^2 \) where the factor \( \hat{\sigma}_B^2 \) is especially utilized to counteract the data scaling problem (as discussed elaborately in [4], [36]). (ii) The \( \frac{1}{N} \) portion of the factor \( \frac{L_B}{N} \) is used to normalize the FIM w.r.t. \( N \). However, note that \( L_B \) (which is absent in EBICR [4], [33]) is also included as part of the normalizing term because for the mean-squared-error of \( \hat{\sigma}^2 \) to approach the Cramer-Rao bound, we require that \( NL \gg K_B L_B L \) or in other words \( N/L_B \gg K_B \). Hence, we use the normalization factor \( L_B/N \) instead of just \( 1/N \) in (15).

In this way, the penalty will be a function of \( N/L_B \) instead of \( N \) alone (as will be seen in the subsequent steps). This novel modification helps to counteract the effects of changing \( L_B \) on the performance of GEBIC_{R}.

Now, using (3), (10), and (15) we can show that

\[ \left| Q^{-1/2} \hat{F}_I Q^{-1/2} \right| = \frac{L_B}{\hat{\sigma}_B} \frac{\hat{A}_I^T A_I}{N} \frac{I_L}{\frac{L_B}{2\hat{\sigma}_0^2}} \]

\[ = \frac{L_B^k B^L L + \frac{1}{2} L}{2(\hat{\sigma}_B^2) L B L + \frac{1}{2}} \left| I_L \otimes \frac{A_I^T A_I}{N} \frac{L}{N} \right| \]

can be rewritten as

\[ \text{const.} \times \left| I_L \right| \left| \frac{A_I^T A_I}{N} \right| = O(1) \] \hspace{1cm} (16) \]

as \( N \) grows large and/or \( \sigma^2 \to 0 \). Hence, this term can be removed without significantly affecting the criterion. Next, observe that the \( \ln |Q| \) term in 14 can be expanded as follows

\[ \ln |Q| = \ln \left( \frac{N}{L_B} \right) \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right)^{k_B L_B L} + \left( k_B L_B L + 1 \right) \ln \left( \frac{N}{L_B} \right) + \left( k_B L_B L + 2 \right) \ln \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right) . \hspace{1cm} (17) \]

Therefore, using (16) and (17) we can rewrite (14) as

\[ \ln |\hat{F}_I| = k_B L_B L \ln \left( \frac{N}{L_B} \right) + \left( k_B L_B L + 2 \right) \ln \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right) + O(1) + \ln \left( N/L_B \right) . \]

Next, for the model prior probability term \(-2 \ln p(\hat{\theta}_I|\mathcal{H}_I)\) in (13), a similar strategy is adopted as in EBIC [2] such that

\[ \ln p(\hat{\theta}_I|\mathcal{H}_I) \propto \left( p_B \right)^{-\zeta} \] where \( \zeta \geq 0 \) is a tuning parameter. If \( p_B \) is sufficiently large, the following approximation can be assumed \( \ln \left( p_B \right) \approx k_B \ln p_B \). [3]. This gives

\[ -2 \ln p(\hat{\theta}_I|\mathcal{H}_I) = 2 \zeta k_B \ln p_B + \text{const}. \] \hspace{1cm} (19) \]

Now, substituting (18), (19) in (13) and dropping the \( O(1) \), the \( \ln \left( N/L_B \right) \) term (since independent of \( k_B \)), the constant and the \( \ln (N/L_B) \) term we arrive at the GEBIC_{R} of [4], [33].

\[ \text{GEBIC}_{R}(\mathcal{I}) = NL \ln \hat{\sigma}_B^2 + k_B L_B L \ln \left( \frac{N}{2\pi L_B} \right) \]
\[ + \left( k_B L_B L + 2 \right) \ln \left( \frac{\hat{\sigma}_B^2}{\sigma_0^2} \right) + 2 \zeta k_B \ln p_B . \hspace{1cm} (20) \]

In practice, we compute the GEBIC_{R} score block-wise, i.e., \( \text{GEBIC}_{R}(\mathcal{I}_B) \) where \( \mathcal{I}_B \subseteq \{1, \ldots, p_B\} \). Then the \( \hat{\sigma}_B^2_t \) can be replaced by \( \hat{\sigma}_B^2_t = \|\Pi^t(A_I)^tY\|_F^2/NL \). Finally, the true block support is estimated as

\[ \hat{S}_B = \arg \min_{\mathcal{I}_B} \left\{ \text{GEBIC}_{R}(\mathcal{I}_B) \right\} \] \hspace{1cm} (21) \]

Observe that for \( L = L_B = 1 \), GEBIC_{R} boils down to EBIC_{R} [4]. Thus GEBIC_{R} can cater to all forms of linear regression scenarios.

**IV. Simulation Results**

In this section, we provide numerical simulations to highlight the performance of GEBIC_{R} for model selection in BMMV models. We consider the linear model \( Y = AX + W \), where the design matrix \( A \) is generated with independent entries following normal distribution \( N(0, 1) \). The cardinality of the true block-support \( S_B \) is chosen to be \( K_B = 4 \). Also,
without loss of generality, we assume \( S_B = [1, 2, 3, 4] \). The non-zero entries in \( X \) are randomly assigned \( \pm 1 \). The SNR in dB = \( 10 \log_{10} (\sigma_s^2/\sigma^2) \), where \( \sigma_s^2 \) and \( \sigma^2 \) denote signal and true noise power, respectively. The signal power is computed as \( \sigma_s^2 = \| AX \|_F^2/NL \). The chosen SNR (dB) and \( \sigma^2 \) are then used to determine the noise power as \( \sigma^2 = \sigma_s^2/10^{(\text{SNR (dB)}/10)} \).

Using this \( \sigma^2 \), the elements of the noise matrix \( W \) are generated following \( W_{i,j} \sim N(0, \sigma^2) \). The probability of correct model selection (PCMS), i.e., \( P_r(S_B = S_{\text{true}}) \) is evaluated over 1000 Monte Carlo trials. At each Monte Carlo trial, a new design matrix \( A \) is generated in order to preserve the randomness in the data. For predictor/subset selection, BMMV-OMP (B-OMP) [27], [39] (Algorithm 1) is utilized because of its ease of use and broad application. BMMV-OMP (B-OMP) [27], [39] (Algorithm 1) is used for predictor/subset selection for its simplicity and wide range of applicability. The performance of GEBIC\(R \) is compared with GRRT and the oracle, which is B-OMP with a-priori knowledge of the block sparsity \( K_B \). Hence, the oracle provides the upper bound on the maximum achievable PCMS for any given setting. The tuning parameters chosen are \( \alpha = 0.01 \) for GRRT (as mentioned in [27]) and \( \zeta = 1 \) (GEBIC\(R \)) [4].

Fig. 1 shows the PCMS vs SNR (dB) with \( N = 150 \) and \( p = 1000 \). Since \( L_B = 10 \), hence, \( p_B = p/L_B = 100 \). Additionally, the performance is shown for two different settings of the \( L \) parameter, viz. \( L = 5 \) and 15 to highlight the influence of \( L \) on the overall behaviour of the methods. The first clear observation is that for the considered tuning parameter setting, both GEBIC\(R \) and GRRT are empirically consistent in high-SNR, i.e., PCMS \( \rightarrow 1 \) as SNR \( \rightarrow \infty \) (or inversely \( \sigma^2 \to 0 \)). Second, compared to GRRT, the performance curve of GEBIC\(R \) is much closer to the oracle, especially for low values of SNR. Furthermore, compared to \( L = 5 \), the oracle plot shifts toward the left when \( L = 15 \). This indicates that increasing \( L \) improves the true support recovery ability of B-OMP, which ultimately improves the model selection performance of the methods.

Fig. 2 presents the PCMS vs \( N \) plot. Here, a fixed value of \( p = 5000 \) is chosen. Additionally, the performance is shown for two different values of the \( L_B \) variable, viz. \( L_B = 5 \) and 20 to highlight the impact of \( L_B \) on the overall model selection performance. A similar trend is observed here as well. Both methods achieve empirical consistency (PCMS \( \rightarrow 1 \)) as \( N \) grows large. However, GEBIC\(R \) provides slightly better performance compared to GRRT for smaller \( N \) values, and is much closer to the oracle performance. Furthermore, we also observe that increasing \( L_B \) lowers the support recovery performance of B-OMP, which is evident from the shift in the oracle performance towards the right. Thus, requiring more samples to achieve the same PCMS for \( L_B = 20 \) as compared to \( L_B = 5 \). This ultimately lowers the overall performance of all model selection methods. Hence, we can say that increasing \( L_B \) affects the performance negatively.

![Fig. 1. PCMS vs SNR (dB) for N = 150, p = 1000, L = [5, 15], L_B = 10 and K_B = 4.](image1)

![Fig. 2. PCMS vs N for SNR = -4 dB, p = 5000, L = 5, L_B = [5, 20] and K_B = 4.](image2)
V. CONCLUSION

In this paper, we presented GEBIC$_R$ which is a generalized version of EBIC$_R$ to handle model selection in the block-sparse HD general linear regression. GEBIC$_R$ is applicable to all forms of the linear regression structure such as SMV, BSMV, MMV, and BMMV thus making it a versatile IC. The steps to arrive at the criterion are shown in detail. Simulation results show that GEBIC$_R$ is an empirically consistent criterion as $N \to \infty$ and/or SNR $\to \infty$. Also, its performance for lower SNR and $N$ values is close to the oracle behaviour. Furthermore, we also underline the manner in which the parameters $L$ and the block length $L_B$ affect the overall model selection performance.

REFERENCES