# Co-factor analysis of citation networks*

Alex Hayes†
Department of Statistics, University of Wisconsin-Madison
and
Karl Rohe
Department of Statistics, University of Wisconsin-Madison

October 3, 2024

**Abstract**

One compelling use of citation networks is to characterize papers by their relationships to the surrounding literature. We propose a method to characterize papers by embedding them into two distinct "co-factor" spaces: one describing how papers send citations, and the other describing how papers receive citations. This approach presents several challenges. First, older documents cannot cite newer documents, and thus it is not clear that co-factors are even identifiable. We resolve this challenge by developing a co-factor model for asymmetric adjacency matrices with missing lower triangles and showing that identification is possible. We then frame estimation as a matrix completion problem and develop a specialized implementation of matrix completion because prior implementations are memory bound in our setting. Simulations show that our estimator has promising finite sample properties, and that naive approaches fail to recover latent co-factor structure. We leverage our estimator to investigate 255,780 papers published in statistics journals from 1898 to 2024, resulting in the most comprehensive topic model of the statistics literature to date. We find interpretable co-factors corresponding to many statistical subfields, including time series, variable selection, spatial methods, graphical models, GLM(M)s, causal inference, multiple testing, quantile regression, semiparametrics, dimension reduction, and several more.

*Keywords:* co-factor models, spectral network analysis, matrix completion, missing data, stochastic blockmodels

*The Version of Record of this manuscript has been published and is available in the Journal of Computational Graphics and Statistics at `http://www.tandfonline.com/10.1080/10618600.2024.2394464`.

†This project was supported by the NSF under Grant DMS-1916378; NSF under Grant DMS-1612456; and the ARO under Grant W911NF-15-1-0423.

# 1 Introduction

Suppose we have a collection of written documents, and these documents cite each other. For example, the documents might be academic papers, judicial opinions, or patents, among other possibilities. One useful way to understand individual documents in the collection, and the collection as a whole, is to find documents that cite, and are cited, in similar ways. These documents are likely to be about the same subject, and can thus reveal information about important topics in the corpus.

We develop a network-based approach to understanding the structure in citation corpora, called `CitationImpute`. `CitationImpute` begins by representing a corpus as a network, where each document corresponds to a node, and citations between documents correspond to directed edges. Then, it uses a spectral factorization technique to embed each document into two distinct latent spaces, one characterizing how papers cite, and the other characterizing how papers get cited.

Unlike prior approaches to citation analysis, `CitationImpute` models citations from older documents to newer documents as structurally missing. As a consequence, our algorithm must estimate singular subspaces via matrix completion methods. Existing matrix completion methods are computationally prohibitive in this setting, so we develop a singular subspace estimator with reasonable time and space complexity.

After estimating singular subspaces, `CitationImpute` uses varimax rotation to identify latent factors in the network (as opposed to k-means, or k-medians clustering). This allows each document to have a weighted membership in each cluster. The overall procedure can be understood intuitively in the context of stochastic blockmodels, but is appropriate for a much broader class of low-rank network models.

We validate the new procedure with a simulation study, finding that the new estimator recovers latent factors under a partially observed stochastic blockmodel. Finally, we analyze 255,780 statistics papers and 2.2 million citations published in journals on statistics and probability, producing a comprehensive breakdown of topics in the statistics literature. We present the keywords most associated with these topics in Table 1 (factors describing how papers get cited) and Table 2 (factors describing how papers cite).

`CitationImpute` is related to several lines of extant work, most notably empirical investigations of the academic statistics literature. Selby (2020) and Stigler (1994) consider relationships between statistics papers and the larger academic literature, with Selby (2020) reviewing approaches to community detection in networks and suggesting a number of diagnostic techniques for assessing model fit. Ji et al. (2022), an expansion of Ji and Jin (2016), considers a dataset with about a third as many papers as our own, and investigates undirected (and dynamic) networks of academic authors based on co-authorship and co-citation. Ji et al. (2022) estimates researcher interests by embedding researchers into a three-dimensional latent space. In contrast, we model the topics of individual manuscripts, and co-embed manuscripts into much more detailed thirty-dimensional "sending" and "receiving" latent spaces. Rohe and Zeng (2023) co-factor a directed network of journal-journal citation counts using varimax factor analysis, but aggregates citations over time and thus avoids the chronological missingness we consider here.

Table 1: Keywords for Y (incoming citation) factors

| Factor Name | Top words | ID |
|---|---|---|
| non-convex penalties | selection, variable, penalized, oracle, lasso, nonconcave | y01 |
| feature screening | screening, dimensional, ultrahigh, feature, independence, high | y02 |
| bayesian model selection | bayesian, models, complexity, disease, model, fit | y03 |
| post-selection inference | high, dimensional, lasso, regression, confidence, dantzig | y04 |
| survival analysis | survival, censored, hazards, proportional, cox, regression | y05 |
| information criteria | model, clustering, mixture, selection, dimension, mixtures | y06 |
| causal inference | propensity, causal, score, observational, treatment, effects | y07 |
| multiple testing | false, discovery, multiple, rate, testing, controlling | y08 |
| graphical models | graphical, covariance, estimation, sparse, lasso, high | y09 |
| bayesian non-parametrics | dirichlet, bayesian, nonparametric, mixture, mixtures, priors | y10 |
| supervised dimension reduction | dimension, reduction, regression, sliced, inverse, sufficient | y11 |
| LASSO (optimization) | lasso, regularization, coordinate, descent, selection, via | y12 |
| LASSO (classic) | lasso, selection, shrinkage, regression, via, longitudinal | y13 |
| kriging | spatial, gaussian, datasets, covariance, large, temporal | y14 |
| empirical likelihood | empirical, likelihood, confidence, ratio, intervals, regions | y15 |
| GLM(M)s | longitudinal, data, generalized, models, estimating, binary | y16 |
| functional data | functional, regression, principal, data, linear, longitudinal | y17 |
| skew normals | skew, normal, distributions, multivariate, distribution, t | y18 |
| quantile regression | quantile, regression, quantiles, censored, median, estimation | y19 |
| bayesian model selection | bayesian, selection, variable, bayes, priors, prior | y20 |
| missing data | missing, imputation, data, longitudinal, nonignorable, nonresponse | y21 |
| adaptive clinical trials | trials, clinical, adaptive, sequential, group, multiple | y22 |
| splines + random effects | models, mixed, splines, smoothing, longitudinal, regression | y23 |
| multivariate analysis | covariance, matrices, high, dimensional, large, matrix | y24 |
| MCMC | monte, carlo, markov, metropolis, chain, bayesian | y25 |
| single index models | coefficient, varying, models, index, single, partially | y26 |
| causal semiparametrics | missing, semiparametric, regression, sampling, data, estimation | y27 |
| individual/optimal treatment | treatment, regimes, individualized, learning, optimal, estimating | y28 |
| RIDGE | ridge, regression, biased, linear, estimators, estimator | y29 |
| cure models | cure, survival, censored, rate, mixture, hazards | y30 |

Methodologically, `CitationImpute` is an extension of the varimax rotation technique studied in Rohe and Zeng (2023), and is closely related to co-clustering methods (Rohe et al., 2016; Choi and Wolfe, 2014; Choi, 2017), as well as clustering methods for bipartite networks (Larremore et al., 2014; Razaee et al., 2019; Yen and Larremore, 2020), some of which can be extended to handle missing data (Zhao et al., 2022; Peixoto, 2018). While there is a large literature on network clustering with missing data, these techniques cannot be used for co-factoring and co-clustering. Nonetheless, some techniques similarly leverage nuclear norm penalized singular subspace estimation to handle missing edges (Chen et al., 2014; Vinayak et al., 2014; Li et al., 2020). There have also been some efforts to incorporate topic structure into preferential attachment models (Pollner et al., 2006; Hajek and Sankagiri, 2019), bridging the gap between mixture modelling and more traditional bibliometric analysis (Price, 1976).

Finally, our work is related to the general matrix completion literature, in particular nuclear norm penalization approaches for estimating partially observed matrices (Kim and Choi, 2013; Gu et al., 2014; Klopp, 2014; Cui et al., 2015; Hosono et al., 2016; Gu et al., 2017; Zhang and Ng, 2019; Yang et al., 2022; Shamir and Shalev-Shwartz, 2014; Bhojanapalli and Jain, 2014; Cho et al., 2019; Mazumder et al., 2010). While this literature has recently made impressive inroads regarding the consistency of nuclear-norm regularization for spectral recovery in deterministic and non-uniform sampling settings (Foucart et al., 2021; Zhu et al., 2022), we are unaware of consistency results for the upper triangular observation pattern present in citation data, and thus validate our approach with simulations.

### Notation

Let $\mathbf{u}_i(A), \boldsymbol{\lambda}_i(A), \mathbf{v}_i(A)$ be functions that return the $i^{th}$ left singular vector, singular value, and right singular vector of a matrix $A$, respectively. Similarly, define $\boldsymbol{\lambda}_i^2(A) = (\boldsymbol{\lambda}_i(A))^2$. We use $\langle \cdot, \cdot \rangle$ to denote the Frobenius inner product and $\| \cdot \|_F$ the Frobenius norm. Let $A_{i\cdot}$ denote the $i^{th}$ row of a matrix $A$ and $A_{\cdot j}$ denote the $j^{th}$ column. For a partially observed matrix $A$, let $\Omega_A$ be the set $\{(i, j) : A_{ij} \text{ is observed}\}$ and $\tilde{\Omega}_A$ be the set $\{(i, j) : A_{ij} \text{ is observed and non-zero}\}$; when $A$ is clear from context we will omit the subscript $A$. By $Y_A$ ($Y$ when the context is clear) we denote the binary matrix such that $Y_{ij}$ is one when $(i, j) \in \Omega_A$ and zero otherwise. $\odot$ indicates elementwise multiplication between two matrices with the same dimensions. We use $P_{\Omega_A}(B) = B \odot Y_A$ to denote the projection of a matrix $B$ onto observed support of another matrix $A$, and $P_{\Omega_A}^{\perp}(B) = B \odot (1 - Y_A)$. Let $P_\ell(A)$ denote the "clipping" projection that sets the first $\ell$ columns and the last $\ell$ rows of $A$ all to zero. Finally, $g(n) = \mathcal{O}(f(n))$ means that $\lim_{n\to\infty} g(n)/f(n) \leq M$ for some constant $M$. All proofs are deferred to the Appendix.

# 2 Model

## 2.1 Co-factor model

We use the co-factor model of Rohe and Zeng (2023) as a model for latent similarities between documents. The co-factor model is a low-rank, distributionally agnostic generalization of the stochastic co-blockmodel (Holland et al., 1983; Rohe et al., 2016), and includes sub-models such as stochastic blockmodels, degree-corrected stochastic blockmodels (Karrer and Newman, 2011), (degree-corrected) mixed membership stochastic blockmodels (Airoldi et al., 2008; Jin et al., 2024), latent dirichlet allocation (Blei et al., 2003), and (generalized) random dot product graphs (Lyzinski et al., 2014), many of which are closely related to topic models (Gerlach et al., 2018).

In the co-factor model, each document $i$ possesses two co-factors. One co-factor, $Z_i \in \mathbb{R}^k$, controls outgoing citations, or the topics that a paper is likely to cite, and the other co-factor, $Y_i \in \mathbb{R}^k$, controls incoming citations, or the topics that a paper is likely to be cited by. The co-factor structure of the model operationalizes the fundamental difference between citing and being cited. Mathematically, co-factor models are generalizations of factor models, and there are compelling reasons to model full co-factor structure: co-factor structure is theoretically necessary to capture key features of real world network data (Chanpuriya et al., 2020), an observation empirically verified by Rohe et al. (2016) and Qing and Wang (2022), among others.

**Example 2.1.** Consider Tibshirani (1996), which introduced `LASSO` regression. The `LASSO` paper builds upon a small body of statistical work on variable selection and resampling, but itself forms the basis for a large body of applied work, especially in genomics and biomedical settings. The directionality of citations is clear in the reference counts: Tibshirani (1996) cites twenty papers, but is cited by tens of thousands of papers. If we do not distinguish between papers cited and citing papers, we might fail to distinguish between the genomics literature (incoming co-topic) and the variable selection literature (outgoing co-topic), as well as differing propensities to cite and to be cited.

In the co-factor model, conditional on the latent factors, each edge $A_{ij}$ of the network is sampled independently from a distribution with expectation $\mathcal{A} \equiv \mathbb{E}(A \mid Z, B, Y) = ZBY^T \in \mathbb{R}$ where $B \in \mathbb{R}^{k \times k}$ is a mixing matrix that controls how the outgoing and incoming latent factors interact. In the citation setting, $\mathcal{A}$ represents the similarities between documents in the latent topic space. $B$ is a weighting matrix that describes how likely it is that a document $i$ loading on outgoing factor $Z_{\cdot k}$ forms an edge to a document $j$ loading on incoming factor $Y_{\cdot \ell}$. As the $B$-mediated similarity between the outgoing topic of document $i$ and the incoming similarity of document $j$ increases, (i.e. $\mathcal{A}_{ij}$ gets larger), the probability of citation $i \to j$ goes up.

For the co-factor model to be identified, the co-factors $Z$ and $Y$ and the mixing matrix $B$ must satisfy several assumptions: the mixing matrix $B$ must be full rank, the rows of $Z$ and $Y$ must be independent and identically distributed (that is, $Z_{1\cdot}, Z_{2\cdot}, ..., Z_{n\cdot}$ must be i.i.d, and $Y_{1\cdot}, Y_{2\cdot}, ..., Y_{n\cdot}$ must be i.i.d.), and the distribution of the $Z_{i\cdot}$ and $Y_{i\cdot}$ must be leptokurtic (i.e., skewed). Skewness is the key assumption for $Z$ and $Y$ to be identified. When $Z$ and $Y$ come from leptokurtic distributions, the co-factors $Z$ and $Y$ are identified up to sign-flips and permutations of the column order.

The co-factor model is similar in form to mixed membership stochastic blockmodels, and generalizes the mixed membership stochastic blockmodel (see the supplement of Rohe and Zeng (2023) for a precise characterization). Unlike mixed membership stochastic blockmodels, the rows of $Z$ and $Y$ do not need to normalized, and can take on negative values. In practice, the sign ambiguity of $Z$ and $Y$ can almost always be resolved by forcing the columns of $Z$ and $Y$ to be skew positive, in which case $Z$ and $Y$ typically consist of sparse, axis-aligned, positive values. The sparsity of $Z$ and $Y$ often enables substantive interpretations of the latent factors, as each node typically loads on a small number of factors.

## 2.2 Chronological observation mechanism

To specialize the co-factor model to the citation setting, we incorporate an observation mechanism.

**Definition 2.1.** Given a corpus of documents $i = 1, ..., n$ published at times $T_1, ..., T_n$, the partially observed adjacency matrix is

$$A_{ij} = \begin{cases} 1 & \text{if } T_j \leq T_i \text{ and } i \text{ cites } j, \\ 0 & \text{if } T_j \leq T_i \text{ and } i \text{ does not cite } j, \text{ and} \\ \text{unobserved} & \text{if } T_j > T_i. \end{cases} \tag{2.1}$$

For convenience, we re-index the documents in order of publishing times, forcing $T_1 \geq ... \geq T_n$, such that $T_1$ is the most recent publishing time, and $T_n$ is the earliest publishing time. Using this indexing scheme, the observed portion of the network is nearly upper triangular, but elements can occur in the lower triangle when $T_i = T_j$.

Under the citation observation mechanism, citations from older papers to newer papers are missing. This is because the lack of citations from older papers to newer papers should be uninformative about the outgoing co-factor of the older paper and the incoming co-factor of the newer paper.

If we presume that the older paper definitively cites the newer paper, or definitively does not cite the newer paper, this will force the corresponding co-factors closer together or farther apart in the latent topic space. `CitationImpute` thus treats citations forward-in-time as missing rather than precisely observed zeroes or ones. This allows the estimation procedure to spectrally infer co-factors without introducing chronological artifacts.

**Example 2.2.** Consider Hoerl and Kennard (1970), which introduced `RIDGE` regression. Since the `RIDGE` paper was published long before the `LASSO` paper, Hoerl and Kennard (1970) does not cite Tibshirani (1996). But, since `RIDGE` regression and `LASSO` regression are closely related, it is plausible that the two papers are close to each other in outgoing topic space. The impossibility of citation forward-in-time is uninformative about the latent similarity between the two papers.

**Remark 2.1.** *The chronological observation mechanism is only relevant if citations are directed relationships. If there is no semantic information contained in the direction of a citation, we can impute the lower triangle of $A$ based on the upper triangle of $A$ by setting $A_{ij} = A_{ji}$ for all missing edges.*

**Remark 2.2.** *In some settings, such as the scientific literature, documents might build on each other, with later documents iterating on past work. In a co-factor model, one could argue that this should be modeled as dependence amongst the latent factors. We are not aware of any approaches to handle such dependence, but believe they are an interesting topic for future work.*

## 2.3 Statistical identification of latent co-factors

The chronological observation mechanism presents several challenges. First, it is unclear if the co-factors $Z$ and $Y$ are identified based on the information observed in the upper triangle of $A$.

In Proposition 2.1, we show that outgoing community memberships $Z_{i\cdot}$ are identified for all but the very earliest documents, and that the incoming community memberships $Y_{i\cdot}$ are identified for all but the most recent documents. Some co-factors are unidentified because the most recent documents have not been around long enough to possibly be cited by papers from all topics and because oldest documents were written too early to possibly cite papers from all topics.

More precisely, Proposition 2.1 states that if the conditional expectation of a citation network $\mathcal{A}$ is rank $k$ and the $\ell_{\text{z}} \times \ell_{\text{y}}$ submatrix in the top right of $\mathcal{A}$ is rank $k$, it is possible to reconstruct all of $\mathcal{A}$ except for the elements in the last $\ell_{\text{z}}$ rows and the elements in the first $\ell_y$ columns. Observing a full rank matrix $M$ in the top right of $\mathcal{A}$ ensures that no information is hidden in the lower triangle (see Figure 1).

The statement of Proposition 2.1 requires some additional notation. Let $\mathscr{R}_{n,k}$ be the set of rank $k$ matrices contained in $\mathbb{R}^{n \times n}$. Imagine that $\mathcal{A}, \mathcal{B} \in \mathscr{R}_{n,k}$ are the conditional expectations

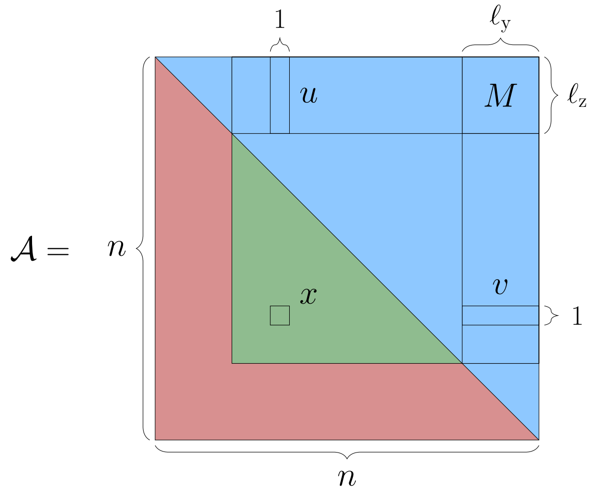


Figure 1: A decomposition of a conditional expectation matrix $\mathcal{A}$. Elements in the upper triangle are observed in the citation setting. We would like to recover elements of the lower triangle based on the information in the upper triangle. When $\text{rank}(\mathcal{A}) = \text{rank}(M)$, a portion of the lower triangle is identified, but the left-most rows and bottom-most columns cannot be recovered.

of two semi-parametric factor models. $P_U(\mathcal{A})$ and $P_U(\mathcal{B})$ are projections of $\mathcal{A}$ and $\mathcal{B}$ onto the space of upper triangular matrices. $P_U(\mathcal{A})$ and $P_U(\mathcal{B})$ represent the conditional expectations of the observed portion of $\mathcal{A}$ and $\mathcal{B}$, respectively.

**Proposition 2.1.** *Let $\mathcal{A}, \mathcal{B} \in \mathscr{R}_{n,k}$. If $P_U(\mathcal{A}) = P_U(\mathcal{B})$ and there exist $\ell_z, \ell_y \in \{k, ..., n/2\}$ such that $M = \mathcal{A}_{[1:\ell_z,(n-\ell_y):n]}$ has rank $k$, then $\mathcal{A}_{ij} = \mathcal{B}_{ij}$ for all $i, j \in \mathbb{Z}$ satisfying $1 < i \leq (n - \ell_z)$ and $\ell_y < j \leq n$.*

**Remark 2.3.** *Proposition 2.1 can be understood constructively as approximating $\mathcal{A}$ from $P_U(A)$ using the Nyström method (Drineas and Mahoney, 2005; Gittens and Mahoney, 2016). Because $\mathcal{A}$ is rank $k$, the Nyström method has zero approximation error.*

When the conditions of Proposition 2.1 are violated, it is possible that no elements in the lower triangle of $\mathcal{A}$ are recoverable. That is, if $M$ is rank $k$ only for $\ell_z, \ell_y > n/2$, there is information hidden in the lower triangle of $\mathcal{A}$ that is lost upon projecting onto the upper triangle. A concrete example where recovery is impossible is given by the following matrix $\mathcal{A}$, which has rank two. Let $J_n$ denote an $n \times n$ matrix of ones and suppose

$$\mathcal{A} = \begin{bmatrix} a\, J_{n/2} & a\, J_{n/2} \\ b\, J_{n/2} & a\, J_{n/2} \end{bmatrix}.$$

Every element in the upper triangle of $\mathcal{A}$ is $a$, and thus there is no way of estimating $b$ when the lower triangle of $\mathcal{A}$ is missing. $\mathcal{A}$ corresponds to a two-block stochastic blockmodel where the first $n/2$ documents are in one block and the last $n/2$ documents are in a separate block. Because these blocks of documents do not overlap in time (here represented by node order) and they have asymmetric citation probabilities, all information about asymmetric citation probabilities is lost. However, if $n$ is large and the rows and columns of $\mathcal{A}$ are permuted according to the same random permutation, then there are $\ell_{\text{z}}, \ell_{\text{y}} \ll n/2$ that satisfies the conditions of Proposition 2.1 with high probability.

For balanced stochastic co-blockmodels, such as the one used in the simulation study, it is sufficient to take $\ell_{\text{z}} = \ell_{\text{y}} = 2k \log k$ to achieve identifiability with high probability. This demonstrates that identification with $\ell_{\text{z}}, \ell_{\text{y}} \ll n/2$ is reasonable in blockmodel-like settings.

**Proposition 2.2.** *Suppose $\mathcal{A}$ characterizes the expected adjacency matrix of the simulation test-bed model (Definition 4.2). Let $\ell_z = \ell_y = 2k \log k$ and let $M$ be as defined in Proposition 2.1. Then* $\mathbb{P}(\text{rank}(M) = k) = 1 - 2n^{-1}$.

# 3 Spectral estimation

Spectral clustering typically proceeds in three steps. First, the network is represented as a matrix, often the adjacency matrix, but sometimes normalized or regularized versions of the graph Laplacian. Second, the leading singular vectors of this matrix are estimated, which associates each node in the graph with a point in Euclidean space. Lastly, these node embeddings are analyzed using standard methods for Euclidean data. While this estimation strategy may seem ad hoc, spectral estimators performs statistical inference under network models that are identified by their singular subspaces, a large class of models that includes stochastic blockmodels and many generalizations thereof (von Luxburg, 2007; Ji and Jin, 2016; Jin, 2015; Lei and Rinaldo, 2015; Rohe et al., 2016; Lyzinski et al., 2017; Athreya et al., 2015, 2018; Lyzinski et al., 2014; Priebe et al., 2019).

## 3.1 The algorithm

`CitationImpute` adapts the standard spectral estimation pipeline to the citation setting. The main difference is that we cannot estimate singular subspaces of the adjacency matrix using a singular value decomposition, due to missing data. Instead, `CitationImpute` uses the `AdaptiveImpute` algorithm of Cho et al. (2019), which is a self-tuning variant of the `softImpute` algorithm of Mazumder et al. (2010).

`CitationImpute` accepts as input a network adjacency matrix $A \in \mathbb{R}^{n \times n}$ where the lower triangle is assumed to be mostly missing, a desired number of co-factors $k \in \{2, ..., n\}$, and clipping parameters $\ell_{\text{z}}, \ell_{\text{y}} \in \{k, ..., n/2\}$. The algorithm then proceeds as follows.

1. Set all elements in the first $\ell_{\text{y}}$ columns of $A$ and last $\ell_{\text{z}}$ rows of $A$ to zero. This means that edges corresponding to the unidentified rows of $Z$ and $Y$ are ignored during estimation; see Proposition 2.1 for details.

2. Estimate the singular vectors and singular values of $A \approx \widehat{U}\widehat{D}\widehat{V}^T$ using `AdaptiveImpute` (Cho et al., 2019). In Section 3.2 we describe why a naive implementation is computationally infeasible, and in Section 3.3 we outline our computational contributions and a practical implementation of `AdaptiveImpute` for upper triangular data.

3. Compute the varimax rotations of $\widehat{U}$ and $\widehat{V}$ and construct rotated singular vector matrices $\widehat{Z}, \widehat{B}$ and $\widehat{Y}$, respectively. We briefly review varimax rotation in Section 3.4.

`CitationImpute` has several hyperparameters: the number of desired co-factors $k$, and the clipping parameters $\ell_z$ and $\ell_y$. In simulations, we find that $\ell_z = \ell_y = n/10$ are good default values, but recommend applying domain knowledge as appropriate, and conducting a sensitivity analysis (see Section 5.2 for an example).

## 3.2 The computational problem

One contribution of this paper is a collection of algebraic identities (Propositions 3.1 and 3.2) that allow for an efficient implementation of `AdaptiveImpute` on citation matrices with hundreds of thousands of documents.

To understand why these identities are useful, we must disambiguate between two senses of sparsity. A matrix is *sparse* if most of its elements are zero. These matrices can be represented very efficiently on a computer by recording only the small number of non-zero elements and their indices. On the other hand, a matrix is *sparsely observed* if only a few of its entries are observed, regardless of the value of those entries. These two notations of sparsity are often conflated, and sparsely observed matrices are often represented as sparse matrices, where implicit zeroes are considered missing, and the observed zeroes must be explicitly tracked.

In the citation setting, the data matrix $A$, as defined in (2.1), is densely observed; at least half of the entries are defined by the data. However, in the portion of the network that is observed, the data is sparse, i.e., mostly zero-valued. Thus, the usual conflation of sparse and sparsely observed matrices leads to issues: there are $n\,(n-1)/2$ elements in the upper triangle of $A$ that must be explicitly tracked even if they are zero. Using this representation, even moderately sized corpora cannot be held in memory on commodity hardware. Beyond memory considerations, adding approximately $n\,(n-1)/2$ explicit zeroes to a sparse matrix slows down matrix operations like matrix-vector multiplication.

This makes matrix completion algorithms infeasible in both time and space when using the naive sparse representation of $A$. Both `AdaptiveImpute` and `softImpute` rely on iterated singular value decompositions of a running low-rank approximation $\tilde{A}^{(t)}$ to $A$. In the typical setting where the number of nodes is $n$, the rank of the decomposition is $k$, and $n \gg k$, naively taking a singular value decomposition of $\tilde{A}^{(t)}$ has time complexity per iteration $\mathcal{O}\left(n^2\,k\right)$. This high computational complexity constrains researchers to inference on networks with at most thousands of nodes.

We are able to reduce the both the time and space complexity of the matrix completion problem. The solution requires leveraging the fact that $A$ is sparse, even if it is not sparsely observed. In particular, there is no need to explicitly track zeroes in the upper triangle of $A$, and $A$ may be represented as a sparse matrix that records only non-zero elements of $A$ and zeroes in the lower triangle of $A$. Using this representation, with some algebraic tricks, all the operations necessary for `AdaptiveImpute` are computationally feasible. In brief, by representing $\tilde{A}^{(t)}$ as the sum of four carefully constructed matrices, we can reduce the naive time complexity from $\mathcal{O}(n^2\,k)$ down to $\mathcal{O}(|\tilde{\Omega}|\,k + n\,k^2)$, where $|\tilde{\Omega}|$ is the number of observed non-zero elements of $A$. In real world datasets $|\tilde{\Omega}|$ represents the number of citations between documents, and empirical evidence suggests that each document in a citation network cites a fixed number of other documents, regardless of the overall size of the corpus. That is, $|\tilde{\Omega}|$ is $\mathcal{O}(n)$. Thus the effective per-iteration runtime reduces from $\mathcal{O}(n^2\,k)$ to $\mathcal{O}(n\,k^2)$.

## 3.3 `AdaptiveImpute`

The `AdaptiveImpute` algorithm is similar to `softImpute` (Hastie et al., 2015), with two key differences. First, `AdaptiveImpute` initializes with a debiased singular value decomposition. Second, on each iteration, `AdaptiveImpute` adaptively varies the `softImpute` thresholding parameter. This procedure is defined in Algorithm 1, which is identical to the algorithm as defined in (Cho et al., 2019) but with some minor notation changes and the introduction of a maximum number of iterations $T$.

---

**Algorithm 1:** AdaptiveImpute

---

**Input:** partially observed matrix $A \in \mathbb{R}^{n \times n}$, rank $k \in \{2, ..., n\}$, convergence tolerance $\varepsilon > 0$, and maximum allowable iterations $T \in \mathbb{Z}^+$.

$Z^{(1)} \leftarrow$ `AdaptiveInitialize`$(A, k)$

**repeat**

$\tilde{A}^{(t)} \leftarrow P_\Omega(A) + P_\Omega^\perp\left(Z^{(t)}\right)$

$\widehat{V}_i^{(t)} \leftarrow \mathbf{v}_i\left(\tilde{A}^{(t)}\right)$ for $i = 1, ..., k$

$\widehat{U}_i^{(t)} \leftarrow \mathbf{u}_i\left(\tilde{A}^{(t)}\right)$ for $i = 1, ..., k$

$\tilde{\alpha}^{(t)} \leftarrow \dfrac{1}{n-k} \sum_{i=k+1}^{n} \boldsymbol{\lambda}_i^2\left(\tilde{A}^{(t)}\right)$

$\hat{\lambda}_i^{(t)} \leftarrow \sqrt{\boldsymbol{\lambda}_i^2\left(\tilde{A}^{(t)}\right) - \tilde{\alpha}^{(t)}}$ for $i = 1, ..., k$

$Z^{(t+1)} \leftarrow \sum_{i=1}^{k} \hat{\lambda}_i^{(t)} \widehat{U}_i^{(t)} \widehat{V}_i^{(t)T}$

$t \leftarrow t + 1$

**until** $\|Z^{(t+1)} - Z^{(t)}\|_F^2 / \|Z^{(t+1)}\|_F^2 < \varepsilon$ *or* $t \geq T$

**return** $\hat{\lambda}_i^{(t)}, \widehat{U}_i^{(t)}, \widehat{V}_i^{(t)}$ *for* $i = 1, ..., k$

---

The initializer is given by running Algorithm 3, which we defer to the appendix. If we compute $Z^{(1)}$ by taking a rank $k$ singular value decomposition of $P_\Omega(A)$ and fix $\alpha^{(t)} = \lambda$ for all $t$ (note that $\tilde{\alpha}^{(t)}$ is the data adaptive thresholding parameter), `AdaptiveImpute` reduces to `softImpute`. This implies that a naive implementation of `AdaptiveImpute` inherits the per-iteration time complexity of `softImpute`, which is $\mathcal{O}(|\Omega|\, k + n\, k^2)$, plus the cost of evaluating $\tilde{\alpha}^{(t)}$.

### 3.3.1 Feasible implementation

In practice, the runtime for each iteration of `AdaptiveImpute` and `softImpute` is dominated by the singular value decomposition, which is computed using an algorithm such as the implicitly restarted Lanczos bidiagonalization algorithm. The time complexity of this decomposition depends fundamentally on an underlying bidiagonalization subroutine (Algorithm 2), and the time complexity of the bidiagonalization subroutine in turn depends on cost of left and right matrix-multiplication of $\widetilde{A}^{(t)}$ with an appropriately sized vector (Baglama and Reichel, 2005).

When $A$ is sparsely observed, $\tilde{A}^{(t)}$ can be expressed as a sparse matrix plus a low-rank matrix

$$\tilde{A}^{(t)} = \underbrace{P_\Omega(A - Z^{(t)})}_{\text{sparse}} + \underbrace{Z^{(t)}}_{\text{low-rank}}, \quad \text{sparsely observed setting}$$

and matrix-vector multiplication has time complexity $\mathcal{O}(|\Omega|\, k)$ for the sparse part and $\mathcal{O}(n\, k^2)$ for the low-rank part. In the citation setting, naively re-using this decomposition in the bidiagonalization subroutine is inefficient since $|\Omega| \approx n^2/2$.

However, a similar trick can improve the time complexity of multiplication with $\tilde{A}^{(t)}$: we can drop observed zeroes from consideration if we partition $\tilde{A}^{(t)}$ carefully. Since $P_{\tilde{\Omega}}(A) = P_\Omega(A)$, we can compute only on $\tilde{\Omega}$. Let $U = \{(i, j) : i < j\}$ denote the indices of the upper triangle of $A$ and $L$ denote the indices of the observed elements of $A$ on the lower triangle, such that $\Omega = U \cup L$. Then

$$\begin{aligned}
\tilde{A}^{(t)} &= P_\Omega(A) + P_\Omega^\perp(Z^{(t)}) \qquad \text{citation setting} \\
&= P_\Omega(A) - P_\Omega(Z^{(t)}) + P_\Omega(Z^{(t)}) + P_\Omega^\perp(Z^{(t)}) \\
&= P_{\tilde{\Omega}}(A) - P_\Omega(Z^{(t)}) + Z^{(t)} \\
&= \underbrace{P_{\tilde{\Omega}}(A)}_{\text{sparse}} - \underbrace{P_L(Z^{(t)})}_{\text{sparse}} - \underbrace{P_U(Z^{(t)})}_{\text{low-rank until projection}} + \underbrace{Z^{(t)}}_{\text{low-rank}}.
\end{aligned}$$

Efficient implementation strategies for matrix-vector multiplications with the sparse and low-rank terms are well known. This leaves the $P_U(Z^{(t)})$ term, which is low-rank until it is projected onto the upper triangle. There one can use the same implementation strategy as for the low-rank component, but summing over fewer indices.

**Proposition 3.1.** *Let $Z^{(t)} \in \mathbb{R}^{n \times n}$ be a rank $k$ matrix with singular value decomposition $Z^{(t)} = UDV^T$ and let $x \in \mathbb{R}^n$. Then*

$$\left[P_U(Z^{(t)})\, x\right]_i = \langle U_{i\cdot}, \tilde{W}_i \rangle,$$

*where $\tilde{W}_{ki} = \sum_{j=i+1}^n W_{kj}$ and $W_{\cdot j} = (DV^T)_{\cdot j} \cdot x_j$.*

We defer the proof to the appendix. Proposition 3.1 is a straightforward result that suggests a computational scheme for evaluating the term $P_U(Z^{(t)})\, x$. In particular, it suggests constructing $W$, then $\tilde{W}$, and then obtaining elements of $P_U(Z^{(t)})\, x$ element by element. This procedure requires $\mathcal{O}(n\, k^2)$ flops as opposed to the $\mathcal{O}(n^2\, k)$ flops of a naive implementation. The left-multiplication case is analogous.

The last requirement to implement `AdaptiveImpute` is a similarly efficient calculation of $\alpha^{(t)}$.

**Proposition 3.2.** *Let $\tilde{A}^{(t)}, Z^{(t)}$ and $\alpha^{(t)}$ be as defined in Algorithm 1. Recall that $Z^{(t)}$ is a low-rank matrix of the form $UDV^T$ with $U, V \in \mathbb{R}^{n \times k}$ orthonormal and $D \in \mathbb{R}^{k \times k}$ diagonal. Then*

$$\alpha^{(t)} = \frac{1}{n-k} \left[ \|P_{\tilde{\Omega}}(A)\|_F^2 + \left\|Z^{(t)}\right\|_F^2 - \left\|P_L\left(Z^{(t)}\right)\right\|_F^2 - \left\|P_U\left(Z^{(t)}\right)\right\|_F^2 - \sum_{i=1}^k \boldsymbol{\lambda}_i^2\left(\tilde{A}^{(t)}\right) \right].$$

*Additionally, define* $U^{rq} \in \mathbb{R}^n$ *and* $V^{rq\triangle} \in \mathbb{R}^n$ *such that*

$$U_i^{rq} = U_{ir}\, U_{iq}, \qquad \text{and} \qquad V_i^{rq\triangle} = \sum_{j=i+1}^{n} (DV)_{rj}^T (DV)_{qj}^T \qquad \forall i = 1, ..., n.$$

*Then*

$$\left\| P_U \left( Z^{(t)} \right) \right\|_F^2 = \sum_{r=1}^{k} \sum_{q=1}^{k} \left\langle U^{rq}, V^{rq\triangle} \right\rangle.$$

To understand the computational complexity of this expression we proceed term by term. First, consider the $\sum_{i=1}^k \boldsymbol{\lambda}_i^2(\tilde{A}^{(t)})$ term. Each iteration of `AdaptiveImpute` computes a truncated singular value decomposition of $\tilde{A}^{(t)}$ of rank $k$ before computing $\alpha^{(t)}$, so evaluating this term is a trivial $\mathcal{O}(k)$ summation since $\boldsymbol{\lambda}_i(\tilde{A}^{(t)})$ is available for $i = 1, ..., k$. Next, observe that $\|P_{\tilde{\Omega}}(A)\|_F^2$ and $\|P_L(Z^{(t)})\|_F^2$ are collectively $\mathcal{O}(|\tilde{\Omega}|\, k)$. This leaves the terms $\|Z^{(t)}\|_F^2$ and $\|P_U(Z^{(t)})\|_F^2$, both of which can require $\mathcal{O}(n\, k^2)$ flops. As in Proposition 3.1, the idea is that evaluating $\|P_U(Z^{(t)})\|_F^2$ is essentially the same evaluating $\|Z^{(t)}\|_F^2$ case, modulo some care while indexing. The time complexity to compute $\alpha^{(t)}$ is then $\mathcal{O}(|\tilde{\Omega}|\, k + n\, k^2)$ flops. Using this scheme to evaluate $\alpha^{(t)}$, the overall time complexity of each iteration of `AdaptiveImpute` is $\mathcal{O}(|\tilde{\Omega}|\, k + n\, k^2)$.

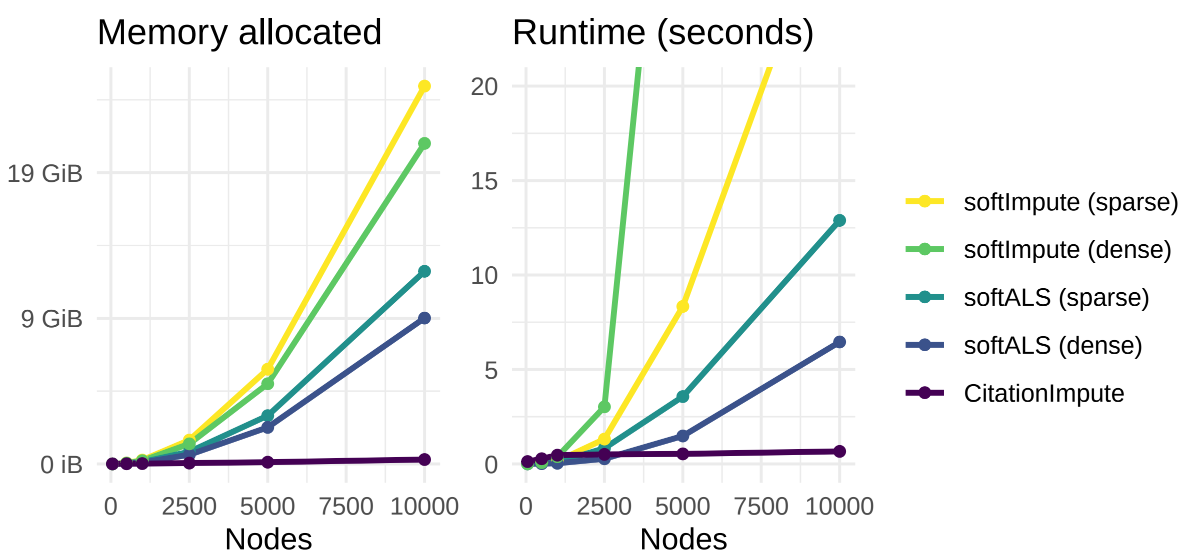


Figure 2: Comparison of memory and time complexity of `CitationImpute` with some existing options for low-rank matrix completion when applied our simulation test-bed model (Definition 4.2). Some run-times are truncated at 20 seconds in the right panel. Each estimator is iterative; we compare time and memory use for five iterations. Existing implementations for both sparse and dense data representations are memory bound and do not scale to networks with more than several thousand nodes. Our implementation, although un-optimized, uses less memory and is faster for large networks.

### 3.4 Varimax rotation

After obtaining an estimated singular value decomposition $A \approx \widehat{U}\widehat{D}\widehat{V}^T$ from `AdaptiveImpute`, `CitationImpute` varimax rotates the estimates to obtain latent factors for each node. Given an $n \times k$ matrix orthonormal matrix $U$, varimax rotation finds a $k \times k$ orthogonal matrix $R$ that maximizes

$$v(R, U) = \sum_{\ell=1}^{k} \frac{1}{n} \sum_{i=1}^{n} \left( [UR]_{i\ell}^4 - \left( \frac{1}{n} \sum_{j=1}^{n} [UR]_{j\ell}^2 \right)^2 \right)$$

over the set of $k \times k$ orthonormal matrices. In particular, it compute $\widehat{R}_U$ that maximizes $v(\cdot, \widehat{U})$ and $\widehat{R}_V$ that maximizes $v(\cdot, \widehat{V})$ where $\widehat{R}_U$ and $\widehat{R}_V$ are $k \times k$ orthonormal matrices. Calculating these rotation matrices is a routine operation available in many statistical packages. After the rotation matrices $\widehat{R}_U$ and $\widehat{R}_V$ have been found, the latent factors are estimated as

$$\widehat{Z} = \sqrt{n}\widehat{U}R_{\widehat{U}}, \quad \widehat{Y} = \sqrt{n}\widehat{V}R_{\widehat{V}}, \quad \text{and} \quad \widehat{B} = R_{\widehat{U}}^T \widehat{D} R_{\widehat{V}}/n. \tag{3.1}$$

Rohe and Zeng (2023) show that, for $\widehat{U}, \widehat{V}$ obtained from the singular value decomposition in the fully observed case, varimax rotated estimates $\widehat{Z}, \widehat{B}$ and $\widehat{Y}$ are consistent for population terms $Z, B$ and $Y$.

## 4 Simulation study

To assess the performance of `CitationImpute`, we perform a simulation study using a co-stochastic blockmodel, a sub-model of the co-factor model. In the simulation study, `CitationImpute` recovers singular subspaces of $\mathcal{A}$ and the latent factors $Z$ and $Y$ at the same rate as an oracle estimator that has access to all of $A$. The simulations also show that naive imputation of missing data leads to inconsistent estimates.

For the simulations, we use a Poisson degree-corrected stochastic co-blockmodel subject to lower triangular missingness.

**Definition 4.1** (Degree-corrected stochastic co-blockmodel)**.** The *degree-corrected stochastic co-blockmodel* is random graph model on $n$ nodes. Each node $i$ is assigned an incoming community $z(i) \in \{1, ..., k\}$ and an outgoing community $y(i) \in \{1, ..., k\}$ according to parameters $\pi^{\text{in}} \in [0,1]^k$ and $\pi^{\text{out}} \in [0,1]^k$, such that $\mathbb{P}(z(i) = j) = \pi_j^{\text{in}}$ and $\mathbb{P}(y(i) = j) = \pi_j^{\text{out}}$ for $j \in \{1, ..., k\}$. Each node $i$ is also assigned a propensity $\theta_i^{\text{out}} \in \mathbb{R}_+$ to send edges, and a propensity $\theta^{\text{in}} \in \mathbb{R}_+$ to receive edges. Conditional on community memberships and edge formation propensities, integer-valued edges occur independently according to a Poisson distribution with expectation.

$$\mathbb{E}(A_{ij} \mid z(i), y(j)) = \theta_i^{\text{out}} \, B_{z(i),y(j)} \, \theta_j^{\text{in}}.$$

where $B \in [0,1]^{k \times k}$ is a rank $k$ mixing matrix denoting propensities of edge formation between communities. $B$ can be rescaled by a constant to enforce that the expected density of edges in the network is $\rho$.

The idea behind the simulation model is to mimic the behavior we expect in citation networks, where papers in a given field will primarily cite papers from that same field (strong diagonal structure in $B$), but will intermittently cite papers from other fields (some active elements of $B$ on the off-diagonal). This is motivated by the observation that the topics that Tibshirani (1996) cites and the topics that cite Tibshirani (1996) are distinct.

**Definition 4.2** (simulation model)**.** The simulation model is a degree-corrected stochastic co-blockmodel with $n$ nodes, $k$ co-communities, and expected density $\rho = 0.15$. Let $\pi_j^{\text{in}} = \pi_j^{\text{out}} = 1/k$ for $j = 1, ..., k$, such that the co-communities are balanced. Let be $\theta^{\text{in}}$ and $\theta^{\text{out}}$ be generated by sampling $n$ independent realizations from an exponential distribution with mean eight, and then adding one to each realization, inducing some degree-heterogeneity. The diagonal elements of $B$ are set to $B_{\text{within}} = 0.8$. $k$ elements of the off-diagonal to $B_{\text{between}} = (B_{\text{within}}/3 - (k-2)\, B_{\text{inactive}})$ (in particular, the off-diagonal values in the first row of $B$, and the last element of the second column of $B$). The remaining elements of the off diagonal to $B_{\text{inactive}} = 0.01$. This ensures that $B$ is rank $k$ and that there is strong assortative structure in the network. In the simulations, we use $k \in \{3, 6, 9\}$, with corresponding values of $B_{\text{between}} = 0.257, 0.227, 0.197$.

We compare the `CitationImpute` to an oracle estimator with access to the full data $A$, and also two imputation estimators. In total, we compare four estimators:

1. `CitationImpute`, with $\ell_{\text{z}} = \ell_{\text{y}} = n/10$,
2. singular value decomposition applied after imputing all missing data as zeros (call this the *zero-imputed* estimator),
3. singular value decomposition applied after imputing all missing data by symmetrizing the observed data (call this the *symmetrized* estimator), and
4. oracle singular value decomposition applied to a fully observed similarity data (call this the *fully observed* estimator).

For the last three estimators, after estimating singular subspace, the singular vectors are varimax rotated according to (3.1) to obtain co-factor estimates.

To measure how well various estimators recover the singular subspaces of $\mathcal{A}$, we compute the $\sin\Theta$ distance between the subspaces spanned by $U$ and $\widehat{U}$ (Vu and Lei, 2013; Bhatia, 1997), for identified rows only. Given two orthonormal bases $U \in \mathbb{R}^{n \times k}$ and $\widehat{U} \in \mathbb{R}^{n \times k}$, the singular values $\sigma_1, ..., \sigma_k$ of $U^T \widehat{U}$ are the cosines of the principal angles $\cos\theta_1, ..., \cos\theta_k$ between the span of $U$ and the span of $\widehat{U}$. Define $\sin\Theta(U, \widehat{U})$ to be a diagonal matrix containing the sine of the principle angles of $U^T \widehat{U}$. Then the $\sin\Theta$ distance between the subspaces spanned by $U$ and $\widehat{U}$ is given by

$$d(U, \widehat{U}) = \| \sin\Theta(U, \widehat{U}) \|_F.$$

We aggregate error across identified rows of the estimates $\widehat{U}$ and $\widehat{V}$ and report a single metric

$$\mathcal{L}_{\text{subspace}}(U, \widehat{U}, V, \widehat{V}) = \| \sin\Theta(U, \widehat{U}) \|_F + \| \sin\Theta(V, \widehat{V}) \|_F.$$

To measure how well the estimators recover the latent factors $Z$ and $Y$, we report root mean squared error on individual elements of identified rows $\widehat{Z}$ and $\widehat{Y}$. Since varimax estimates $\widehat{Z}$ and $\widehat{Y}$ are only determined up to sign-flips and column reordering, this requires an alignment step to match $\widehat{Z}$ with $Z$, and $\widehat{Y}$ with $Y$. Let $\mathscr{P}(k)$ be the set of $k \times k$ orthogonal matrices whose entries $P_{ij}$ are elements of $\{-1, 0, 1\}$. Define

$$P_Z = \arg\min_{P \in \mathscr{P}(k)} \|Z - \widehat{Z}P\|_F \tag{4.1}$$

$$P_Y = \arg\min_{P \in \mathscr{P}(k)} \|Y - \widehat{Y}P\|_F. \tag{4.2}$$

We find $P_Z$ and $P_Y$ by using the Hungarian algorithm to match columns of the estimates $\widehat{Z}, \widehat{Y}$ to the corresponding population values $Z, Y$. Then the elementwise factor root mean squared error is

$$\mathcal{L}_{\text{factor}}(Z, \widehat{Z}, Y, \widehat{Y}) = \sqrt{\frac{1}{n\,k}\left(\|Z - \widehat{Z}P_Z\|_F^2 + \|Y - \widehat{Y}P_Y\|_F^2\right)}.$$

To perform the simulation, we evaluate the subspace loss and the factor loss 200 times for every estimator, every $k \in \{3, 6, 9\}$, and every $n \in \{100, 182, 331, 603, 1099, 2000\}$. In Figure 3, we report the average subspace loss and the average factor loss for these combinations. Estimation error for `CitationImpute` decreases at approximately $\sqrt{n}$-rates, suggesting that `CitationImpute` is a consistent estimator of the singular subspaces of $\mathcal{A}$ and also of the latent factors $Z$ and $Y$. The rate for `CitationImpute` parallels that of the oracle estimator with access to all of $A$, although it unsurprisingly advantageous to observe the full data.

In contrast, the symmetric imputation strategy and the zero-imputation strategies are not reliable ways to estimate singular subspaces or latent factors. Estimation error for both imputation strategies is constant as a function of $n$, suggesting that estimators based on naive imputation approaches are inconsistent. The symmetric imputation strategy is always better than treating the unobserved entries as zeroes, which makes sense as the model has some underlying symmetry. Some additional simulation results investigating the imputation estimators are available in Appendix B.

# 5 Analysis of the statistics literature

We next leveraged `CitationImpute` to analyze of the academic statistics literature.

## 5.1 Data

We used proprietary Web of Science data that we obtained through an institutional agreement with Clarivate Analytics. The complete Web of Science corpus contains hundreds of millions of documents, which amount to nearly a terabyte of data. We considered only papers published in a subset of 125 journals focused on probability and statistics (see Appendix D for a list of the journals). The node-induced subgraph formed by considering only these papers and the citations between them had 281,883 nodes, 2,224,775 edges, and $24,051$ weakly connected components (a weakly connected component in a subgraph where there is a path between every pair of nodes, ignoring the direction of edges). Most of the 24,051 weakly connected

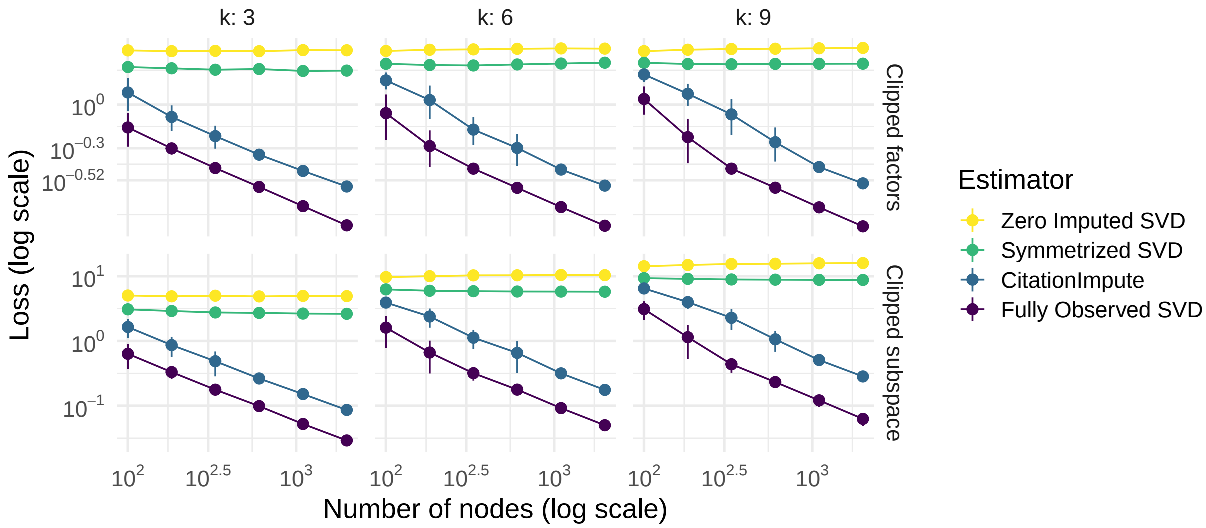


Figure 3: Average estimation error as a function of sample size, on log − log scale. The top row of panels visualizes estimation error of the factors $Z$ and $Y$, excluding clipped factors. The bottom row of panels visualizes estimation error of the singular subspaces, again excluding clipped rows of $U$ and $V$. Each column of panels represents a simulation model with a different number of latent communities. Within each panel, each line corresponds to the loss of a single estimator. Average loss plus and minus one standard deviation are shown as a dotplot; in most cases the standard deviations are too small to see.

components were singletons. The largest weakly connected component contained 255,780 nodes and 2,222,363 edges. From this point onward, when we refer to the "citation network" or "citation graph" we are referring exclusively to this largest connected component. For each document we additionally knew the authors, publication date, and the abstract text, although some of this information was missing.

Papers in the citation network were published between 1898 and 2024. The number of citations received from other papers in the largest connected component (i.e. in-degree) ranged from 0 to 4,759 and the number of citations sent to other papers in the largest connected component (i.e. out-degree) ranged from 0 to 603. There are several articles in the sample that cite hundreds of other papers; these articles are typically bibliographies or reviews. A small number of papers mutually cited each other.

## 5.2 Methods

First we constructed the partially observed adjacency matrix of the citation graph. We ordered nodes chronologically, and then clipped data using $\ell_z = 100,000$ and $\ell_y = 50,000$. This amounted to discarding outgoing citations for papers published before 2004, and incoming citations for papers published after 2018. These clipping parameters were primarily selected on the basis of domain knowledge – we supposed that the topics in the modern statistics literature were present by 2004. We did not estimate incoming co-factors for papers published after 2018, because it can take several years to publish an academic paper, and we believed papers in 2018 are the latest papers that reasonably had the chance to be discovered, cited, and included in the Web of Science dataset.

We then ran `AdaptiveImpute` to obtain a low-rank decomposition $A \approx \widehat{U}\widehat{D}\widehat{V}^T$. Here we report the results for a rank $k = 30$ decomposition. After computing a low-rank decomposition $A \approx \widehat{U}\widehat{D}\widehat{V}^T$, we performed varimax rotation of $\widehat{U}$ and $\widehat{V}$ to obtain a final low-rank decomposition $A \approx \widehat{Z}\widehat{B}\widehat{Y}^T$, as described in Section 3. The rows of $\widehat{Z}$ and the rows of $\widehat{Y}$ thus correspond to document-level latent co-factors (Rohe and Zeng, 2023; Rohe et al., 2016). The rows of $\widehat{Z}$ contained outgoing-citation factors, and the rows of $\widehat{Y}$ contained incoming-citation factors. Both $\widehat{Z}$ and $\widehat{Y}$ were relatively sparse. To interpret the co-factors $\widehat{Y}$ and $\widehat{Z}$, we took several approaches.

First, we found keywords most associated with each factor by examining the words in paper titles following the "best features" approach of Zhang et al. (2021) and Chen (2021). We constructed a document-term matrix from the manuscript title. Letting $X \in \mathbb{Z}^{255{,}780 \times 11298}$, $X_{i\ell}$ indicates the number of times word $\ell$ appears in manuscript title $i$. We restricted our analysis to words that appeared in at least five manuscript titles. Then, for each factor $j$, define the sets $in(j) = \{i : \widehat{Y}_{ij} \geq 0\}$ and $out(j) = \{i : \widehat{Y}_{ij} < 0\}$. Then the importance of word $\ell$ to factor $j$ is

$$\texttt{bff}(j, \ell) = \sqrt{\frac{\sum_{i \in in(j)} \widehat{Y}_{ij} X_{i\ell}}{\sum_{j \in in(i)} \widehat{Y}_{ij}}} - \sqrt{\frac{\sum_{i \in out(j)} X_{i\ell}}{|out(j)|}},$$

and in Tables 1 and 2 we report the six words most important to each factor. To complement this keyword analysis, we found the papers with the largest loadings for each dimension of $\widehat{Y}$ and $\widehat{Z}$, which we refer to as hub papers (see Table 6 and 7 in the Appendix).

## 5.3 Results

The incoming and outgoing co-factors were interpretable and associated with meaningful statistical sub-fields. We found co-factors corresponding to statistical sub-field such as GLMM(s), GEE, multiple testing, feature selection, post-selection inference, survival analysis, MCMC, causal inference, clinical trial design, experimental design, functional data, multivariate analysis, graphical models, semiparametrics, kriging, model selection (both Bayesian and frequentist).

One particularly interesting feature of the $\widehat{Y}$ co-factors was the presence of numerous incoming dimensions related to penalized regression. These factors covered the `LASSO` proper (y13), optimization methods for $L_1$ penalization (y14), non-convex penalties (y01), post-selection inference (y04), feature screening (y02), graphical models (y09) and RIDGE regression (y29). Several other incoming $\widehat{Y}$ co-factors were interesting because they corresponded to more niche statistical subfields. For example, we found incoming factors corresponding to empirical likelihood (y15), supervised dimension reduction (y11), and skew normals (y18). We suspect these co-factors emerged due to strong assortative structure in the sub-field: that is, a tendency to cite heavily within the factor while citing limited papers outside the factor. The tendency for spectral methods to find assortative clusters is widely known within spectral clustering literature, and it makes sense that they would pick up smaller but self-contained topics.

Table 2: Keywords for Z (outgoing citation) factors

| Factor Name | Top words | ID |
|---|---|---|
| non-convex penalties | selection, variable, dimensional, high, penalized, lasso | z01 |
| experimental design | screening, dimensional, high, ultrahigh, feature, supersaturated | z02 |
| bayesian spatial stats | bayesian, models, spatial, model, longitudinal, hierarchical | z03 |
| post-selection inference | high, dimensional, lasso, recurrent, selection, regression | z04 |
| survival analysis | survival, hazards, censored, cox, data, proportional | z05 |
| mixture models | selection, clustering, model, mixture, models, mixtures | z06 |
| causal inference | propensity, causal, score, treatment, missing, observational | z07 |
| multiple testing | false, discovery, testing, multiple, rate, microarray | z08 |
| graphical models | graphical, high, dimensional, models, sparse, estimation | z09 |
| bayesian non-parametrics | bayesian, dirichlet, nonparametric, mixture, clustering, process | z10 |
| supervised dimension reduction | dimension, reduction, sufficient, index, inverse, sliced | z11 |
| times series | garch, volatility, series, models, time, change | z12 |
| sparse multivariate analysis | selection, lasso, high, sparse, variable, dimensional | z13 |
| kriging | spatial, spatio, temporal, gaussian, fields, bayesian | z14 |
| empirical likelihood | empirical, likelihood, inference, missing, partially, jackknife | z15 |
| GEE | longitudinal, data, generalized, binary, estimating, clustered | z16 |
| functional data | functional, data, regression, longitudinal, principal, linear | z17 |
| skew normals | skew, normal, distributions, multivariate, distribution, t | z18 |
| quantile regression | quantile, regression, quantiles, censored, composite, expectile | z19 |
| bayesian model selection | bayesian, selection, variable, priors, prior, model | z20 |
| missing data | missing, imputation, data, longitudinal, with, nonignorable | z21 |
| adaptive clinical trials | adaptive, trials, clinical, sequential, designs, group | z22 |
| splines + random effects | models, mixed, splines, penalized, regression, additive | z23 |
| multivariate analysis | high, dimensional, covariance, matrices, matrix, factor | z24 |
| MCMC | bayesian, carlo, monte, mcmc, metropolis, chain | z25 |
| single index models | varying, coefficient, models, index, single, partially | z26 |
| joint longitudinal/survival models | longitudinal, mixed, models, data, joint, effects | z27 |
| causal inference reviews | causal, treatment, effects, propensity, instrumental, effect | z28 |
| RIDGE | ridge, regression, estimator, liu, linear, estimators | z29 |
| cure models | cure, censored, survival, model, rate, data | z30 |

Most of the incoming co-factors $\widehat{Y}$ correspond closely with an outgoing co-factor $\widehat{Z}$ on the same topic. For instance, there is an incoming survival analysis co-factor (y05) and also an outgoing survival analysis co-factor (z05). The hubs for the incoming co-factor are highly cited methods papers such as Cox (1972) and Andersen and Gill (1982). The hubs for the outgoing co-factor are review papers that cite many of these works while reviewing few citations themselves, such as Guo and Zeng (2014) and Kalbfleisch and Schaubel (2023). To investigate correspondences between $\widehat{Y}$ and $\widehat{Z}$ factors, we plotted the mixing matrix $\widehat{B}$ in the left panel of Figure 4.

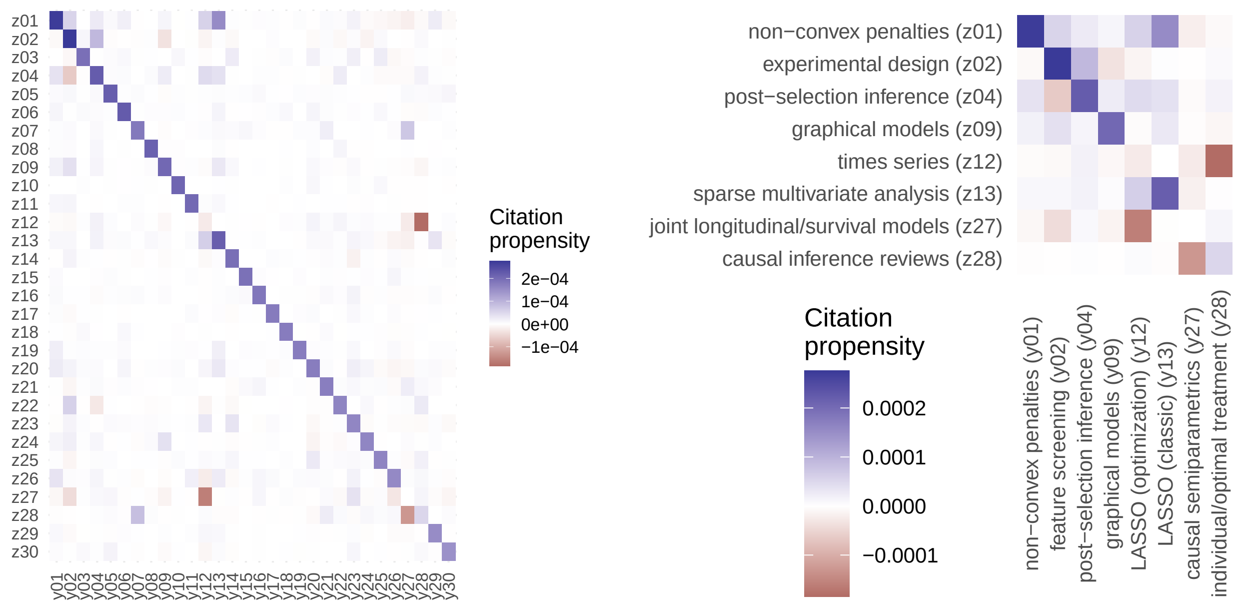


Figure 4: Left: The varimax estimate $\widehat{B}$. Each entry $\widehat{B}_{ij}$ denotes the estimated citation propensity from papers loading on $i^{th}$ outgoing co-factor $Z_{\cdot i}$ to the $j^{th}$ incoming co-factor $Y_{\cdot j}$. Right: A labelled sub-matrix of $\widehat{B}$ considering the co-factors exhibiting off-diagonal structure.

We found several $\widehat{Z}$ factors that did not correspond closely with any incoming $\widehat{Y}$ factor. For instance, z12 is a co-factor describing propensity to cite papers on times series analysis, while y12 is a co-factor related to the LASSO and optimization. Similarly, z27, a co-factor about joint longitudinal models, and z28, a causal inference co-factor, did not exhibit topical correspondence with y27, on causal semiparametrics, and y28, on individualized treatment rules. We visualized the relationships between the unmatched factors in the right panel of Figure 4, where it is clear that some co-factors are not in one-to-one correspondence with one another.

One question was how to interpret co-factors exhibiting one-to-one incoming-to-outgoing correspondence. For example, what was the difference between the outgoing survival analysis factor (z10) and the incoming survival analysis factor (y10)? To answer this question, we looked at the hub papers for each co-factor. For the survival analysis factor, for example, the top incoming hub was Cox (1972), which introduced the proportional hazards model, and the top outgoing hub was Guo and Zeng (2014), a survey of semiparametric models in survival analysis. Incoming $\widehat{Y}$ hub papers were typically highly cited, important papers in each sub-field. In contrast, the outgoing $\widehat{Z}$ hub papers were typically review articles, retrospectives, tutorials, and papers with good literature reviews that summarized the past literature. Put differently, statistical papers tended to either: (1) perform important synthesis of past work but be cited very little, (2) cite a limited number of papers while receiving many citations, or (3) cite and be cited very little.

This distinct behavior from the $\widehat{Y}$ paper hubs and $\widehat{Z}$ paper hubs is evidence of co-factor structure in the statistics literature, and more broadly, evidence that papers do indeed cite and get cited in fundamentally different ways.

### 5.4 Sensitivity to choice of rank and clipping parameters

We repeated our analysis for $k \in \{5, 10, 20, 30, 40\}$, holding $\ell_{\text{z}} = 100,000$ and $\ell_{\text{y}} = 50,000$ fixed. Factor keywords, hubs, and mixing matrices for these analyses can be found in the supplemental material. We obtained qualitatively consistent results across values of $k$, and found that the co-factors can be coherently interpreted at all values of $k$ that we explored. In practice, increasing $k$ revealed additional, finer-grain factor structure. We chose to analyze $k = 30$ co-factors because those co-factors revealed rich structure in the statistics literature while remaining interpretable and digestible.

We additionally explored $\ell_{\text{z}} = \ell_y \in \{1, 25000, 50000, 70000\}$, holding $k = 30$. Factors keywords, hubs, and mixing matrices for these analyses can also be found in the supplemental material. We found that the $Y$ keywords, the $Y$ factor hubs, and the $Z$ factor hubs remained fairly stable across choices of $\ell_{\text{z}}$ and $\ell_{\text{y}}$. However, $Z$ keywords and factor identities were more varied, and mixing matrices $\widehat{B}$ also exhibited substantial variation. At low clipping parameters, $\widehat{B}$ exhibited substantial off-diagonal structure. As the clipping parameters increased, $\widehat{B}$ became more and more diagonal. The results for $\ell_{\text{z}} = 100,000$ and $\ell_{\text{y}} = 50,000$ had the least off-diagonal structure in $\widehat{B}$.

Altogether, the sensitivity analysis for the clipping parameters indicated that $\widehat{B}$, and to a lesser extent, the outgoing co-factors $Z$, were somewhat unstable across hyperparameter values. Ultimately, our choice of $\ell_{\text{z}} = 100,000$ and $\ell_{\text{y}} = 50,000$ was based on domain knowledge: we assumed that it would take until 2004 for all outgoing co-topics to appear in the statistical literature, and that papers published after 2018 would not have the chance to be cited by papers from each incoming co-topic, due to the lengthy academic publication process. Regardless, since we do not definitely know how to select $\ell_{\text{z}}$ and $\ell_{\text{y}}$, results should be treated as somewhat tentative.

### 5.5 How the past would cite the future

One of the interesting features of our missing data framework is that it allows us to impute latent similarities from older documents to newer documents, or, with conceptual abuse, citations forward in time. In particular, if a paper $i$ was published before paper $j$, we can estimate the latent similarity from paper $i$ to paper $j$ via the real-valued imputation $\widehat{A}_{ij} \approx \widehat{Z}_{i\cdot} \, \widehat{B} \, \widehat{Y_{j\cdot}}^T$. We suggest interpreting these imputed similarities as you would interpret probability estimates from a linear probability model; as in the linear probability model, we have no guarantee that $\widehat{A}_{ij} \in [0, 1]$, such that $\widehat{A}_{ij}$ represents a valid probability of "citation".

However, we can still think of $\widehat{A}_{ij}$ as indicative of probability of citation, had citation been possible.

In particular, for each paper, we calculated all of these imputed similarities from prior papers. Summing over these imputations, we obtained an estimate of the number of times papers from the past would have cited papers from the future on the basis of topical similarity, were they so able. We computed these estimates for each of the papers in our citation network and report the 15 papers with the highest imputed in-degree in Table 3 and the 15 papers with highest imputed out-degree in Table 4. Most of the papers with high imputed in-degree are related to feature screening, the graphical `LASSO`, or some form of high dimensional regression. Most of the papers with high imputed out-degree are review articles published in Biometrika.

Table 3: Imputed incoming citations (identified edges only)

| Title | Imputed | Cited by |
|---|---|---|
| On asymptotically optimal confidence regions and tests for high-dimensional models (2014) | 1632 | 360 |
| Confidence intervals for low dimensional parameters in high dimensional linear models (2014) | 1564 | 350 |
| Sure independence screening for ultrahigh dimensional feature space (2008) | 1387 | 905 |
| Estimating individualized treatment rules using outcome weighted learning (2012) | 1215 | 280 |
| Regularization paths for generalized linear models via coordinate descent (2010) | 1135 | 1124 |
| Feature screening via distance correlation learning (2012) | 1094 | 327 |
| A robust method for estimating optimal treatment regimes (2012) | 1014 | 210 |
| Model-free feature screening for ultrahigh-dimensional data (2011) | 871 | 246 |
| Sure independence screening in generalized linear models with np-dimensionality (2010) | 847 | 305 |
| Double/debiased machine learning for treatment and structural parameters (2018) | 831 | 222 |
| Nonparametric independence screening in sparse ultra-high-dimensional additive models (2011) | 812 | 262 |
| Exact post-selection inference, with application to the lasso (2016) | 757 | 188 |
| Performance guarantees for individualized treatment rules (2011) | 753 | 219 |
| Simultaneous analysis of lasso and dantzig selector (2009) | 728 | 617 |
| Sparse inverse covariance estimation with the graphical lasso (2008) | 717 | 754 |

Table 4: Imputed outgoing citations (identified edges only)

| Title | Imputed | Cites |
|---|---|---|
| Bayesian statistics in medicine: a 25 year review (2006) | 754 | 511 |
| Joint modeling of longitudinal and time-to-event data: an overview (2004) | 172 | 36 |
| Joint longitudinal-survival-cure models and their application to prostate cancer (2004) | 165 | 34 |
| Methodological issues with adaptation of clinical trial design (2006) | 155 | 41 |
| Adaptive statistical analysis following sample size modification based on interim review of effect size (2005) | 143 | 26 |
| Semiparametric regression during 2003-2007 (2009) | 137 | 219 |
| A 25-year review of sequential methodology in clinical studies (2007) | 135 | 85 |
| Group sequential and adaptive designs - a review of basic concepts and points of discussion (2008) | 134 | 76 |
| Maximum likelihood estimation in semiparametric regression models with censored data (2007) | 131 | 53 |
| Adaptive seamless designs: selection and prospective testing of hypotheses (2007) | 128 | 61 |
| A regulatory view on adaptive/flexible clinical trial design (2006) | 128 | 30 |
| An overview of statistical approaches for adaptive designs and design modifications (2006) | 124 | 35 |
| An investigation of two-stage tests (2006) | 122 | 30 |
| Efficient group sequential designs when there are several effect sizes under consideration (2006) | 122 | 28 |
| Joint modeling of longitudinal and survival data via a common frailty (2004) | 118 | 21 |

# 6 Discussion

We proposed a new method to co-factor documents in citation networks. The method is motivated by the observation that factors should be based on similarity measurements, and citations are only partially observed similarity measurements. Factoring a partially observed network complicated standard spectral clustering procedures and required use of matrix completion methods to estimate singular subspaces of the graph adjacency matrix. Here we found computational difficulties due to the precise observation pattern of citation data, which we resolved via a careful new implementation of the `AdaptiveImpute` algorithm. Because of dependence in the observation mechanism in the citation setting, existing theoretical results for `AdaptiveImpute`, and nuclear norm minimization more generally, were not applicable, and we validated our approach to matrix completion via a simulation study.

Our work suggests several avenues for methodological and theoretical exploration. Methodologically, it may be interesting to propose computationally efficient estimation procedures for other matrix completion methods in the upper triangular observation setting, or more generally in settings where sparse data is densely observed. Methods designed for independent but general sampling distributions, such as weighted nuclear norm minimization, may perform particularly well in the citation setting. Alternatively, further computational improvements would allow for larger scale bibliometric exploration of scientific citation networks. Current bibliometric databases contain hundreds of millions of papers and billions of references, more data than our method can handle. While our analysis of the statistics literature is one of the most extensive to date, incorporating additional papers could illuminate the relationships between statistical methodology and scientific practice at large. Another open question is how to extend our approach to the tensor, or multi-layer, citation network case, which would be appropriate for data like U.S. Court Opinions, where there are several distinct and explicitly labelled types of citation that documents may use when referencing each other. Finally, it may be of significant practical use to develop a better theoretical understanding of how matrix completion methods perform in settings with dependent observation mechanisms.


# Acknowledgements

We thank Steve Meyer at UW-Madison Libraries for assistance with the Web of Science dataset; Keith Levin, Vivak Patel and several anonymous reviewers for feedback on this manuscript; Yunyi Shen for a code contribution; and Alexander Tahk, Ben Bolker, Mark Padgham, Noam Ross, Max Kuhn, Dan Simpson, Sam Power, Patrick Girardet, and Cannon Lewis for generative discussions throughout the course of the project.


# 7 Replication package

We implemented a proof-of-concept implementation of the `AdaptiveImpute` estimator specialized to the citation setting in the `fastadi` R package, which is available on CRAN and at `https://github.com/RoheLab/fastadi`.

Code to reproduce the simulations and performance comparison is available at `https://github.com/alexpghayes/citation-cofactoring-replication/`. Due to licensing agreements, we cannot publish the Web of Science data. Nonetheless, the replication package contains the code we used to analyze the Web of Science data.

# References

Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008). Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research 9*, 1981–2014.

Andersen, P. K. and R. D. Gill (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics 10*(4), 1100–1120.

Athreya, A., D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, Y. Qin, and D. L. Sussman (2018). Statistical Inference on Random Dot Product Graphs: A Survey. *Journal of Machine Learning Research 18*, 1–92.

Athreya, A., C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman (2015). A Limit Theorem for Scaled Eigenvectors of Random Dot Product Graphs. *Sankhya A: The Indian Journal of Statistics 78*(1), 1–18.

Baglama, J. and L. Reichel (2005). Augmented Implicitly Restarted Lanczos Bidiagonalization Methods. *SIAM Journal on Scientific Computing 27*(1), 19–42.

Bhatia, R. (1997). *Matrix Analysis.* Springer.

Bhojanapalli, S. and P. Jain (2014). Universal Matrix Completion. In *Proceedings of the 31st International Conference on Machine Learning.*

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993–1022.

Chanpuriya, S., C. E. Tsourakakis, C. Musco, and K. Sotiropoulos (2020). Node Embeddings and Exact Low-Rank Representations of Complex Networks. In *34th Conference on Neural Information Processing Systems*, Vancouver, Canada.

Chen, F. (2021). *Spectral Methods for Social Media Data Analysis.* Ph. D. thesis, University of Wisconsin-Madison.

Chen, Y., A. Jalali, S. Sanghavi, and H. Xu (2014). Clustering Partially Observed Graphs via Convex Optimization. *Journal of Machine Learning Research 15*, 2213–2238.

Cho, J., D. Kim, and K. Rohe (2019). Intelligent Initialization and Adaptive Thresholding for Iterative Matrix Completion: Some Statistical and Algorithmic Theory for *Adaptive-Impute*. *Journal of Computational and Graphical Statistics 28*(2), 323–333.

Choi, D. (2017). Co-clustering of nonsmooth graphons. *The Annals of Statistics 45*(No. 4), 1488–1515.

Choi, D. and P. J. Wolfe (2014). Co-clustering separately exchangeable network data. *The Annals of Statistics 42*(1), 29–63.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Cui, Z., D. Zhang, K. Wang, H. Zhang, N. Li, and W. Zuo (2015). Weighted Nuclear Norm Minimization Based Tongue Specular Reflection Removal. *Mathematical Problems in Engineering 2015*, 1–15.

Drineas, P. and M. W. Mahoney (2005). On the Nystrom Method for Approximating a Gram Matrix for Improved Kernel-Based Learning. *Journal of Machine Learning Research 6*, 2153–2175.

Foucart, S., D. Needell, R. Pathak, Y. Plan, and M. Wootters (2021). Weighted Matrix Completion From Non-Random, Non-Uniform Sampling Patterns. *IEEE Transactions on Information Theory 67*(2), 1264–1290.

Gerlach, M., T. P. Peixoto, and E. G. Altmann (2018). A network approach to topic models. *Science Advances 4*(7), 1–11.

Gittens, A. and M. W. Mahoney (2016). Revisiting the Nystrom Method for Improved Large-scale Machine Learning. *Journal of Machine Learning Research 17*, 1–65.

Gu, S., Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang (2017). Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision. *International Journal of Computer Vision 121*(2), 183–208.

Gu, S., L. Zhang, W. Zuo, and X. Feng (2014). Weighted Nuclear Norm Minimization with Application to Image Denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 2862–2869. IEEE.

Guo, S. and D. Zeng (2014). An overview of semiparametric models in survival analysis. *Journal of Statistical Planning and Inference 151–152*, 1–16.

Hajek, B. and S. Sankagiri (2019). Community Recovery in a Preferential Attachment Graph. *IEEE Transactions on Information Theory 65*(11), 6853–6874.

Hastie, T., R. Mazumder, J. D. Lee, and R. Zadeh (2015). Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares. *Journal of Machine Learning Research 16*, 3367–3402.

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics 12*(1), 55–67.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks 5*(2), 109–137.

Hosono, K., S. Ono, and T. Miyata (2016). Weighted tensor nuclear norm minimization for color image denoising. In *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, pp. 3081–3085. IEEE.

Ji, P. and J. Jin (2016). Coauthorship and citation networks for statisticians. *The Annals of Applied Statistics 10*(4), 1779–1812.

Ji, P., J. Jin, Z. T. Ke, and W. Li (2022). Co-citation and Co-authorship Networks of Statisticians. *Journal of Business & Economic Statistics 40*(2), 469–485.

Jin, J. (2015). Fast community detection by SCORE. *The Annals of Statistics 43*(1), 57–89.

Jin, J., Z. T. Ke, and S. Luo (2024). Mixed membership estimation for social networks. *Journal of Econometrics 239*(2), 105369.

Kalbfleisch, J. D. and D. E. Schaubel (2023). Fifty Years of the Cox Model. *Annual Review of Statistics and Its Application 10*(Volume 10, 2023), 1–23.

Karrer, B. and M. E. J. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E 83*(1), 016107.

Kim, Y.-D. and S. Choi (2013). Variational Bayesian View of Weighted Trace Norm Regularization for Matrix Factorization. *IEEE Signal Processing Letters 20*(3), 261–264.

Klopp, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli 20*(1), 282–303.

Larremore, D. B., A. Clauset, and A. Z. Jacobs (2014). Efficiently inferring community structure in bipartite networks. *Physical Review E 90*(1), 012805.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics 43*(1), 215–237.

Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika 107*(2), 257–276.

Lyzinski, V., D. L. Sussman, M. Tang, A. Athreya, and C. E. Priebe (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics 8*(2), 2905–2922.

Lyzinski, V., M. Tang, A. Athreya, Y. Park, and C. E. Priebe (2017). Community Detection and Classification in Hierarchical Stochastic Blockmodels. *IEEE Transactions on Network Science and Engineering 4*(1), 13–26.

Mazumder, R., T. Hastie, and R. Tibshirani (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices. *Journal of Machine Learning Research 11*, 2287–2322.

Mitzenmacher, M. and E. Upfal (2017). *Probability and Computing* (Second edition ed.). Cambridge, United Kingdom ; New York, NY, USA: Cambridge University Press.

Peixoto, T. P. (2018). Reconstructing Networks with Unknown and Heterogeneous Errors. *Physical Review X 8*(4), 041011.

Pollner, P., G. Palla, and T. Vicsek (2006). Preferential attachment of communities: The same principle, but a higher level. *Europhysics Letters (EPL) 73*(3), 478–484.

Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science 27*(5), 292–306.

Priebe, C. E., Y. Park, J. T. Vogelstein, J. M. Conroy, V. Lyzinski, M. Tang, A. Athreya, J. Cape, and E. Bridgeford (2019). On a two-truths phenomenon in spectral graph clustering. *Proceedings of the National Academy of Sciences 116*(13), 5995–6000.

Qing, H. and J. Wang (2022). Directed mixed membership stochastic blockmodel. *arXiv:2101.02307*.

Razaee, Z. S., A. A. Amini, and J. J. Li (2019). Matched Bipartite Block Model with Covariates. *Journal of Machine Learning Research 20*, 1–44.

Rohe, K., T. Qin, and B. Yu (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences 113*(45), 12679–12684.

Rohe, K. and M. Zeng (2023). Vintage factor analysis with Varimax performs statistical inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(4), 1037–1060.

Selby, D. A. (2020). *Statistical Modelling of Citation Networks, Research Influence and Journal Prestige*. Ph. D. thesis, University of Warwick.

Shamir, O. and S. Shalev-Shwartz (2014). Matrix Completion with the Trace Norm: Learning, Bounding, and Transducing. *Journal of Machine Learning Research 15*, 3401–3423.

Stigler, S. M. (1994). Citation Patterns in the Journals of Statistics and Probability. *Statistical Science 9*(1).

Tian, Y. (2004). More on maximal and minimal ranks of Schur complements with applications. *Applied Mathematics and Computation 152*(3), 675–692.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Vinayak, R. K., S. Oymak, and B. Hassibi (2014). Graph Clustering With Missing Data : Convex Algorithms and Analysis. In *Advances in Neural Information Processing Systems*.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing 17*(4), 395–416.

Vu, V. Q. and J. Lei (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics 41*(6), 2905–2947.

Yang, M., Y. Li, and J. Wang (2022). Feature and Nuclear Norm Minimization for Matrix Completion. *IEEE Transactions on Knowledge and Data Engineering 34*(5), 2190–2199.

Yen, T.-C. and D. B. Larremore (2020). Community detection in bipartite networks with stochastic block models. *Physical Review E 102*(3), 032309.

Zhang, X. and M. K. Ng (2019). A Corrected Tensor Nuclear Norm Minimization Method for Noisy Low-Rank Tensor Completion. *SIAM Journal on Imaging Sciences 12*(2), 1231–1273.

Zhang, Y., F. Chen, and K. Rohe (2021). Social Media Public Opinion as Flocks in a Murmuration: Conceptualizing and Measuring Opinion Expression on Social Media. *Journal of Computer-Mediated Communication 27*(1), zmab021.

Zhao, J., M. Sun, F. Chen, and P. Chiu (2022). Understanding Missing Links in Bipartite Networks With MissBiN. *IEEE Transactions on Visualization and Computer Graphics 28*(6), 2457–2469.

Zhu, Z., T. Wang, and R. J. Samworth (2022). High-Dimensional Principal Component Analysis with Heterogeneous Missingness. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(5), 2000–2031.

# A AdaptiveInitialize and Lanczos Bidiagonalization

For convenience, we report the algorithmic details of the `Lanczos Bidiagonalization` and `AdaptiveInitialize` sub-routines.

**Algorithm 2:** Lanczos Bidiagonalization

**Input:** matrix $A \in \mathbb{R}^{\ell \times n}$ or functions for evaluating matrix-vector products with the matrices $A$ and $A^T$, initial vector of unit length $p_1 \in \mathbb{R}^n$, number of bidiagonalization steps $m$

$P_1 \leftarrow p_1$; $q_1 \leftarrow Ap_1$

$\alpha_1 \leftarrow \|q_1\|$, $q_1 \leftarrow q_1/\alpha_1$, $Q_1 \leftarrow q_1$

**for** $j = 1$ *to* $m$ **do**

  $r_j \leftarrow A^T q_1 - \alpha_j p_j$

  **if** $j < m$ **then**

    $\beta_j \leftarrow \|r_j\|$, $p_{j+1} \leftarrow r_j/\beta_j$, $P_{j+1} \leftarrow [P_j, p_{j+1}]$

    $q_{j+1} \leftarrow Ap_{j+1} - \beta_j q_j$

    $\alpha_{j+1} \leftarrow \|q_{j+1}\|$, $q_{j+1} \leftarrow q_{j+1}/\alpha_{j+1}$, $Q_{j+1} \leftarrow [Q_j, q_{j+1}]$

    $B_{j,j+1} \leftarrow \beta_j$

  $B_{jj} \leftarrow \alpha_j$

**return** $P_m, Q_m, B_m, r_m$

`AdaptiveInitialize` is computes an initial estimate for a low-rank decomposition, and was originally reported in Cho et al. (2019). We have made minor notational changes for greater clarity.

Note that the left and right singular vectors estimates are initialized separately in `AdaptiveInitialize`, and $\hat{s}_i \in \{-1, 1\}$ can be used to ensure sign consistency between the singular vector pairs.

# B Additional simulation results

Here, we investigate what happens if one of the naive imputation approaches is correct. We consider two symmetric variants of the simulation model. In the first variant, we symmetrize $\mathbb{E}[\mathcal{A} \mid Z, B, Y]$ by using the asymmetric model from Section 4 and setting $B = (B + B^T)/2$ and $\theta^{\text{in}} = \theta^{\text{out}}$. This yields a directed stochastic blockmodel that is symmetric *in expectation*. We then repeat the simulation study of Section 4 and report the results in Figure 5. In this case, the symmetric imputation strategy is insufficient to estimate $Z$ and $Y$, despite the symmetry in the conditional expectation of $\mathcal{A}$.

Then, we consider a genuinely symmetric generative model, by sampling the upper half of a directed stochastic co-blockmodel, and then setting $A_{ij} = Aji$ for all $i, j \in [n]$. Results

**Algorithm 3:** AdaptiveInitialize

**Input:** partially observed matrix $A \in \mathbb{R}^{n \times n}$, desired rank $k \in \{2, ..., n\}$

$\hat{p} \leftarrow |\Omega_A| / n^2$

$\Sigma_V \leftarrow A^T A - (1 - \hat{p}) \operatorname{diag}\left(A^T A\right)$

$\Sigma_U \leftarrow A A^T - (1 - \hat{p}) \operatorname{diag}\left(A A^T\right)$

$\widehat{V}_i \leftarrow \mathbf{v}_i(\Sigma_V)$ for $i = 1, ..., k$

$\widehat{U}_i \leftarrow \mathbf{u}_i(\Sigma_U)$ for $i = 1, ..., k$

$\tilde{\alpha} \leftarrow \frac{1}{n-k} \sum_{i=k+1}^{n} \boldsymbol{\lambda}_i(\Sigma_V)$

$\hat{\lambda}_i \leftarrow \frac{1}{\hat{p}} \sqrt{\boldsymbol{\lambda}_i(\Sigma_V) - \tilde{\alpha}}$ for $i = 1, ..., k$

$\hat{s}_i \leftarrow \operatorname{sign}\left(\langle \widehat{V}_i, \mathbf{v}_i(A) \rangle\right) \cdot \operatorname{sign}\left(\langle \widehat{U}_i, \mathbf{u}_i(A) \rangle\right)$ for $i = 1, ..., k$

**return** $\hat{s}_i, \hat{\lambda}_i, \widehat{U}_i, \widehat{V}_i$ *for* $i = 1, ..., k$

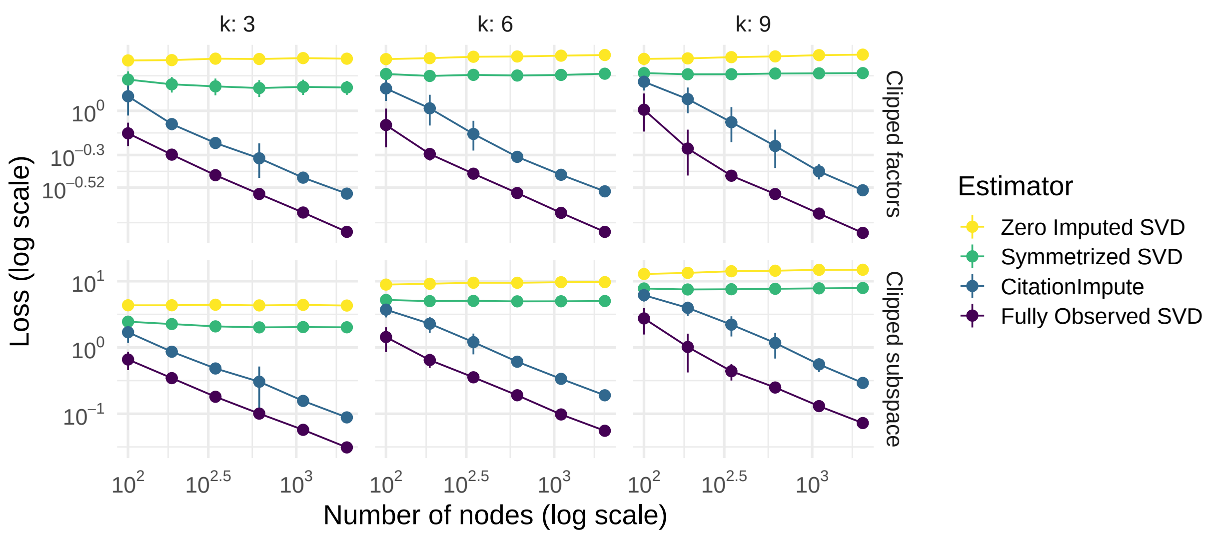


Figure 5: Average estimation error as a function of sample size, on log − log scale. The top row of panels visualizes estimation error of the factors $Z$ and $Y$, excluding clipped factors. The bottom row of panels visualizes estimation error of the singular subspaces, again excluding clipped rows of $U$ and $V$. Each column of panels represents a simulation model with a different number of latent communities. Within each panel, each line corresponds to the loss of a single estimator. Average loss plus and minus one standard deviation are shown as a dotplot; in most cases the standard deviations are too small to see.

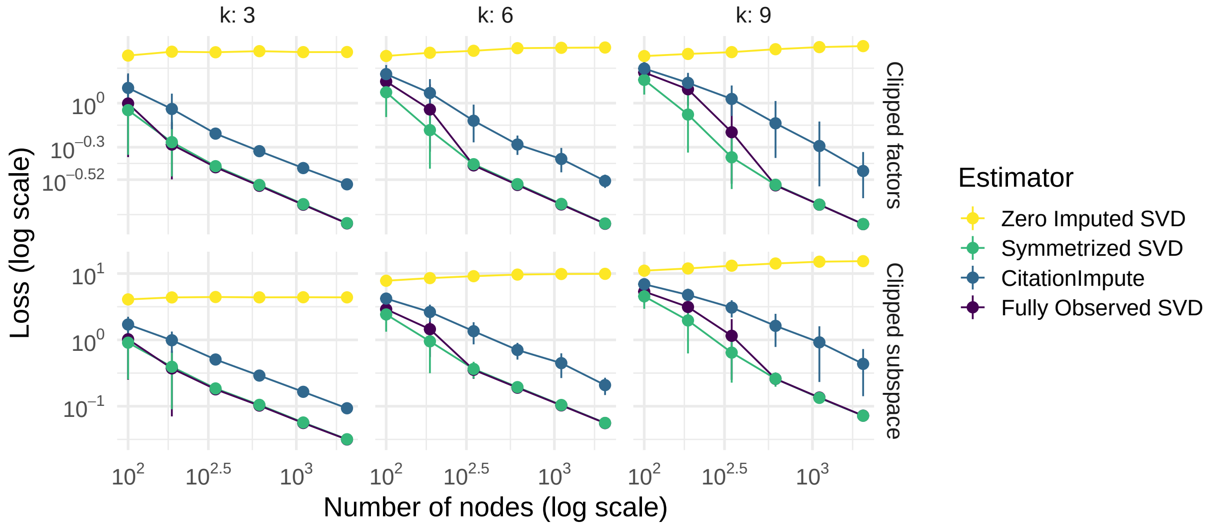


Figure 6: Average estimation error as a function of sample size, on log − log scale. The top row of panels visualizes estimation error of the factors $Z$ and $Y$, excluding clipped factors. The bottom row of panels visualizes estimation error of the singular subspaces, again excluding clipped rows of $U$ and $V$. Each column of panels represents a simulation model with a different number of latent communities. Within each panel, each line corresponds to the loss of a single estimator. Average loss plus and minus one standard deviation are shown as a dotplot; in most cases the standard deviations are too small to see.

for this model are visualized in Figure 6. For truly symmetric $A$, the symmetrized SVD estimator achieves the same performance as the oracle estimator with access to all of $A$. `CitationImpute` continues to achieve the same rate as the oracle and symmetrized estimators in this setting, although with some performance penalty.

# C Proofs

## C.1 Proof of Proposition 2.1

*Proof.* Let $P_U^{-1}(\mathcal{A})$ be the pre-image of $\mathcal{A}$ under $P_U$, which must contain at least one element, and let $\mathcal{C}$ be an arbitrary element of $P_U^{-1}(\mathcal{A})$. Consider a decomposition of $\mathcal{C}$ as in Figure 1. let $i, j \in \mathbb{Z}$ satisfying $1 < i \leq (n - \ell_z)$ and $\ell_y < j \leq n$. That is, $i$ and $j$ index an element of $\mathcal{C}$ that is sent to zero by $P_U$ but that we wish to recover. Put $x \equiv \mathcal{C}_{ij}$. By hypothesis $M = \mathcal{C}_{[1:\ell_z,(n-\ell_y):n]}$. Let $M^-$ be an arbitrary generalized inverse of $M$. From equation (1.5) of Tian (2004) it follows that

$$\text{rank}\left(x - vM^- u\right) \leq \text{rank}\left(\begin{bmatrix} u & M \\ x & v \end{bmatrix}\right) - \text{rank}\left(M\right) = k - k = 0,$$

and thus that $x = vM^- u$. That is, any pre-image of $\mathcal{A}$ is uniquely specified at indexes$i, j \in \mathbb{Z}$ satisfying $1 < i \leq (n - \ell_z)$ and $\ell_y < j \leq n$, as desired. □

## C.2 Proof of Proposition 2.2

*Proof.* In order to show that $M$ has rank $k$, it is sufficient to show that the first $\ell_{\text{z}}$ nodes are collectively members of all $k$ outgoing blocks, and the last $\ell_{\text{y}}$ nodes are collectively members of all $k$ incoming blocks.

The problem thus reduces to the well-known coupon collector's problem. The probability of sampling $k$ out of $k$ distinct and equiprobable items given a sample of size $2k \log k$ is $1 - n^{-1}$ (Mitzenmacher and Upfal, 2017, p125-126). We first use this bound for the incoming $Y$ blocks and then again for the outgoing $Z$ blocks, and combine them with a union bound to complete the proof. □

## C.3 Proof of Proposition 3.1

*Proof.*

$$
\begin{aligned}
\left(P_U(Z_t)\, x\right)_i &= \sum_{j=1}^{n} Z_{ij} \cdot \mathbf{1}(i < j) \cdot x_j \\
&= \sum_{j=1}^{n} \left( \sum_{r=1}^{k} U_{ir} (DV^T)_{rj} \right) \cdot x_j \cdot \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} U_{ir} \sum_{j=1}^{n} (DV^T)_{rj} \cdot x_j \cdot \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} U_{ir} \sum_{j=1}^{n} W_{rj} \cdot \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} U_{ir} \sum_{j=i+1}^{n} W_{rj} \\
&= \sum_{r=1}^{k} U_{ir} \tilde{W}_{ri} \\
&= \langle U_{i\cdot}, \tilde{W}_i \rangle
\end{aligned}
$$

□

## C.4 Proof of Proposition 3.2

*Proof.* First observe that

$$\begin{aligned}
\alpha^{(t)} &= \frac{1}{n-k} \sum_{i=k+1}^{n} \boldsymbol{\lambda}_i^2 \left(\tilde{A}^{(t)}\right) \\
&= \frac{1}{n-k} \left[ \sum_{i=1}^{n} \boldsymbol{\lambda}_i^2 \left(\tilde{A}^{(t)}\right) - \sum_{i=1}^{k} \boldsymbol{\lambda}_i^2 \left(\tilde{A}^{(t)}\right) \right] \\
&= \frac{1}{n-k} \left[ \left\| \tilde{A}^{(t)} \right\|_F^2 - \sum_{i=1}^{k} \boldsymbol{\lambda}_i^2 \left(\tilde{A}^{(t)}\right) \right].
\end{aligned}$$

Further

$$\begin{aligned}
\left\| \tilde{A}^{(t)} \right\|_F^2 &= \left\| P_\Omega(A) + P_\Omega^\perp \left(Z^{(t)}\right) \right\|_F^2 \\
&= \| P_{\tilde{\Omega}}(A) \|_F^2 + \left\| P_\Omega^\perp \left(Z^{(t)}\right) \right\|_F^2 \\
&= \| P_{\tilde{\Omega}}(A) \|_F^2 + \left\| Z^{(t)} \right\|_F^2 - \left\| P_\Omega \left(Z^{(t)}\right) \right\|_F^2 \\
&= \| P_{\tilde{\Omega}}(A) \|_F^2 + \left\| Z^{(t)} \right\|_F^2 - \left\| P_L \left(Z^{(t)}\right) \right\|_F^2 - \left\| P_U \left(Z^{(t)}\right) \right\|_F^2.
\end{aligned}$$

Finally,

$$
\begin{aligned}
\left\| P_U \left( Z^{(t)} \right) \right\|_F^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle U_{i\cdot}, DV_{\cdot j}^T \right\rangle^2 \mathbf{1}(i < j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \sum_{r=1}^{k} U_{ir} (DV)_{rj}^T \right)^2 \mathbf{1}(i < j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \sum_{r=1}^{k} \sum_{q=1}^{k} U_{ir} (DV)_{rj}^T U_{iq} (DV)_{qj}^T \right) \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} U_{ir} (DV)_{rj}^T U_{iq} (DV)_{qj}^T \right) \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left( \sum_{i=1}^{n} U_{ir} U_{iq} \sum_{j=1}^{n} (DV)_{rj}^T (DV)_{qj}^T \right) \mathbf{1}(i < j) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left( \sum_{i=1}^{n} U_{ir} U_{iq} \sum_{j=i+1}^{n} (DV)_{rj}^T (DV)_{qj}^T \right) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left( \sum_{i=1}^{n} U_{ir} U_{iq} \sum_{j=i+1}^{n} (DV)_{rj}^T (DV)_{qj}^T \right) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left( \sum_{i=1}^{n} U_i^{rq} V_i^{rq\triangle} \right) \\
&= \sum_{r=1}^{k} \sum_{q=1}^{k} \left\langle U^{rq}, V^{rq\triangle} \right\rangle,
\end{aligned}
$$

and putting all three of these pieces together completes the proof. □

# D Supplement to data analysis

This section reports the journals included in the data analysis and the factor hubs referenced in the main body of the paper.

Table 5: Journals considered in data analysis

| # | Journal Title |
|---|---|
| 1 | Journal Of Statistical Software |
| 2 | Annual Review Of Statistics And Its Application |
| 3 | Econometrica |
| 4 | Journal Of The American Statistical Association |
| 5 | Journal Of The Royal Statistical Society Series B-Statistical Methodology |
| 6 | Statistical Science |
| 7 | R Journal |
| 8 | Fuzzy Sets And Systems |

Table 5: Journals considered in data analysis *(continued)*

| # | Journal Title |
|---|---|
| 9 | Biostatistics |
| 10 | Statistics And Computing |
| 11 | Ieee-Acm Transactions On Computational Biology And Bioinformatics |
| 12 | Journal Of Business & Economic Statistics |
| 13 | Chemometrics And Intelligent Laboratory Systems |
| 14 | Multivariate Behavioral Research |
| 15 | International Statistical Review |
| 16 | Bayesian Analysis |
| 17 | Annals Of Statistics |
| 18 | Probabilistic Engineering Mechanics |
| 19 | British Journal Of Mathematical & Statistical Psychology |
| 20 | Annals Of Probability |
| 21 | Stochastic Environmental Research And Risk Assessment |
| 22 | Journal Of Computational And Graphical Statistics |
| 23 | Statistical Methods In Medical Research |
| 24 | Quality Technology And Quantitative Management |
| 25 | Journal Of The Royal Statistical Society Series A-Statistics In Society |
| 26 | Econometrics Journal |
| 27 | Probability Theory And Related Fields |
| 28 | Technometrics |
| 29 | Finance And Stochastics |
| 30 | Journal Of Quality Technology |
| 31 | Stata Journal |
| 32 | Open Systems & Information Dynamics |
| 33 | American Statistician |
| 34 | Statistics In Medicine |
| 35 | Biometrics |
| 36 | Scandinavian Actuarial Journal |
| 37 | Annals Of Applied Statistics |
| 38 | Spatial Statistics |
| 39 | Journal Of Agricultural Biological And Environmental Statistics |
| 40 | Journal Of Chemometrics |
| 41 | Biometrika |
| 42 | Advances In Data Analysis And Classification |
| 43 | Journal Of The Royal Statistical Society Series C-Applied Statistics |
| 44 | Annals Of Applied Probability |
| 45 | Bernoulli |
| 46 | Statistical Papers |
| 47 | Statistical Modelling |
| 48 | Biometrical Journal |
| 49 | Stochastic Processes And Their Applications |
| 50 | Statistical Analysis And Data Mining |
| 51 | Stochastics And Partial Differential Equations-Analysis And Computations |
| 52 | Pharmaceutical Statistics |
| 53 | Insurance Mathematics & Economics |
| 54 | Quality Engineering |
| 55 | Annales De L Institut Henri Poincare-Probabilites Et Statistiques |
| 56 | Astin Bulletin |
| 57 | Random Matrices-Theory And Applications |
| 58 | Test |
| 59 | Computational Statistics & Data Analysis |
| 60 | Applied Stochastic Models In Business And Industry |
| 61 | Econometric Theory |
| 62 | Extremes |

Table 5: Journals considered in data analysis *(continued)*

| # | Journal Title |
|---|---|
| 63 | Journal Of Multivariate Analysis |
| 64 | Electronic Journal Of Probability |
| 65 | Journal Of Computational Biology |
| 66 | Environmetrics |
| 67 | Stochastic Analysis And Applications |
| 68 | Journal Of Applied Statistics |
| 69 | Statistica Neerlandica |
| 70 | Environmental And Ecological Statistics |
| 71 | Asta-Advances In Statistical Analysis |
| 72 | Statistica Sinica |
| 73 | Econometric Reviews |
| 74 | Scandinavian Journal Of Statistics |
| 75 | Journal Of Statistical Computation And Simulation |
| 76 | Oxford Bulletin Of Economics And Statistics |
| 77 | Combinatorics Probability & Computing |
| 78 | Electronic Journal Of Statistics |
| 79 | Journal Of Official Statistics |
| 80 | Journal Of Time Series Analysis |
| 81 | Methodology And Computing In Applied Probability |
| 82 | Lifetime Data Analysis |
| 83 | Probability In The Engineering And Informational Sciences |
| 84 | Sort-Statistics And Operations Research Transactions |
| 85 | Statistical Applications In Genetics And Molecular Biology |
| 86 | Stat |
| 87 | Stochastics-An International Journal Of Probability And Stochastic Processes |
| 88 | Annals Of The Institute Of Statistical Mathematics |
| 89 | Computational Statistics |
| 90 | Stochastics And Dynamics |
| 91 | Advances In Applied Probability |
| 92 | Statistics In Biopharmaceutical Research |
| 93 | Journal Of Theoretical Probability |
| 94 | Statistics & Probability Letters |
| 95 | Hacettepe Journal Of Mathematics And Statistics |
| 96 | Journal Of Statistical Planning And Inference |
| 97 | Metrika |
| 98 | Survey Methodology |
| 99 | International Journal Of Biostatistics |
| 100 | Law Probability & Risk |
| 101 | Revstat-Statistical Journal |
| 102 | Canadian Journal Of Statistics-Revue Canadienne De Statistique |
| 103 | Communications In Statistics-Simulation And Computation |
| 104 | Statistics |
| 105 | Brazilian Journal Of Probability And Statistics |
| 106 | Alea-Latin American Journal Of Probability And Mathematical Statistics |
| 107 | Probability And Mathematical Statistics-Poland |
| 108 | Statistical Methods And Applications |
| 109 | Communications In Statistics-Theory And Methods |
| 110 | Journal Of Nonparametric Statistics |
| 111 | Electronic Communications In Probability |
| 112 | Journal Of Biopharmaceutical Statistics |
| 113 | Stochastic Models |
| 114 | Mathematical Population Studies |
| 115 | Journal Of Applied Probability |
| 116 | Sequential Analysis-Design Methods And Applications |

Table 5: Journals considered in data analysis *(continued)*

| # | Journal Title |
|---|---|
| 117 | International Journal Of Game Theory |
| 118 | Infinite Dimensional Analysis Quantum Probability And Related Topics |
| 119 | Journal Of The Korean Statistical Society |
| 120 | Australian & New Zealand Journal Of Statistics |
| 121 | Esaim-Probability And Statistics |
| 122 | Theory Of Probability And Its Applications |
| 123 | Markov Processes And Related Fields |
| 124 | Utilitas Mathematica |
| 125 | Annals Of Mathematical Statistics |

Table 6: Y (incoming citation) factor hubs

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y01 | One-Step Sparse Estimates In Nonconcave Penalized Likelihood Models | 481 | 27 |
| y01 | Tuning Parameter Selectors For The Smoothly Clipped Absolute Deviation Method | 407 | 17 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y02 | Model-Free Feature Screening For Ultrahigh-Dimensional Data | 246 | 17 |
| y02 | Nonparametric Independence Screening In Sparse Ultra-High-Dimensional Additive Models | 262 | 27 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y03 | Winbugs - A Bayesian Modelling Framework: Concepts, Structure, And Extensibility | 645 | 6 |
| y03 | Deviance Information Criteria For Missing Data Models | 311 | 8 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y04 | The Dantzig Selector:: Statistical Estimation When $P$ Is Much Larger Than $N$ | 697 | 11 |
| y04 | Exact Post-Selection Inference, With Application To The Lasso | 188 | 22 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y05 | Partial Likelihood | 933 | 6 |
| y05 | The Lasso Method For Variable Selection In The Cox Model | 398 | 4 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y06 | Model-Based Gaussian And Non-Gaussian Clustering | 429 | 16 |
| y06 | Regression And Time-Series Model Selection In Small Samples | 457 | 12 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |

Table 6: Y (incoming citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y07 | Stratification And Weighting Via The Propensity Score In Estimation Of Causal Treatment Effects: A Comparative Study | 334 | 8 |
| y07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y07 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y08 | Empirical Bayes Analysis Of A Microarray Experiment | 445 | 2 |
| y08 | Strong Control, Conservative Point Estimation And Simultaneous Conservative Consistency Of False Discovery Rates: A Unified Approach | 327 | 10 |
| y09 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y09 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y09 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y09 | A Constrained $L_1$ Minimization Approach To Sparse Precision Matrix Estimation | 320 | 16 |
| y09 | Sparse Permutation Invariant Covariance Estimation | 265 | 22 |
| y10 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y10 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y10 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y10 | Gibbs Sampling Methods For Stick-Breaking Priors | 496 | 30 |
| y10 | Mixtures Of Dirichlet Processes With Applications To Bayesian Nonparametric Problems | 505 | 7 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y11 | On Directional Regression For Dimension Reduction | 263 | 27 |
| y11 | On Principal Hessian Directions For Data Visualization And Dimension Reduction - Another Application Of Steins Lemma | 317 | 29 |
| y12 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y12 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y12 | Least Angle Regression | 1259 | 10 |
| y12 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y12 | Pathwise Coordinate Optimization | 410 | 8 |
| y13 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y13 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y13 | Least Angle Regression | 1259 | 10 |
| y13 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y13 | The Bayesian Lasso | 568 | 11 |
| y14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y14 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y14 | Fixed Rank Kriging For Very Large Spatial Data Sets | 347 | 15 |
| y14 | Hierarchical Nearest-Neighbor Gaussian Process Models For Large Geostatistical Datasets | 190 | 28 |
| y15 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y15 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y15 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y15 | Empirical Likelihood For Linear-Models | 371 | 20 |
| y15 | Methodology And Algorithms Of Empirical Likelihood | 241 | 12 |

Table 6: Y (incoming citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y16 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y16 | Quasi-Likelihood Functions, Generalized Linear-Models, And Gauss-Newton Method | 618 | 3 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y17 | Prediction In Functional Linear Regression | 263 | 20 |
| y17 | Generalized Functional Linear Models | 280 | 20 |
| y18 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y18 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y18 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y18 | Distributions Generated By Perturbation Of Symmetry With Emphasis On A Multivariate Skew $T$ -Distribution | 480 | 17 |
| y18 | A General Class Of Multivariate Skew-Elliptical Distributions | 297 | 3 |
| y19 | Regression Quantiles | 1603 | 21 |
| y19 | Bayesian Quantile Regression | 286 | 7 |
| y19 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y19 | Gibbs Sampling Methods For Bayesian Quantile Regression | 154 | 19 |
| y19 | Goodness Of Fit And Related Inference Processes For Quantile Regression | 210 | 12 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y20 | The Bayesian Lasso | 568 | 11 |
| y20 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y20 | The Horseshoe Estimator For Sparse Signals | 354 | 14 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y21 | Adjusting For Nonignorable Drop-Out Using Semiparametric Nonresponse Models | 480 | 35 |
| y21 | Pattern-Mixture Models For Multivariate Incomplete Data | 352 | 14 |
| y21 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y22 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y22 | Evaluation Of Experiments With Adaptive Interim Analyses | 277 | 7 |
| y22 | Adaptive Sample Size Calculations In Group Sequential Trials | 241 | 16 |
| y22 | A Simple Sequentially Rejective Multiple Test Procedure | 748 | 3 |
| y22 | Multiple Testing Procedure For Clinical-Trials | 621 | 1 |
| y23 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y23 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y23 | Selecting The Number Of Knots For Penalized Splines | 314 | 10 |
| y24 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y24 | Covariance Regularization By Thresholding | 399 | 20 |
| y24 | Regularized Estimation Of Large Covariance Matrices | 392 | 14 |
| y24 | Large Covariance Estimation By Thresholding Principal Orthogonal Complements | 214 | 63 |
| y24 | Determining The Number Of Factors In Approximate Factor Models | 434 | 5 |

Table 6: Y (incoming citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y25 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y25 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y25 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y25 | Markov-Chains For Exploring Posterior Distributions | 924 | 11 |
| y25 | The Calculation Of Posterior Distributions By Data Augmentation | 1035 | 5 |
| y26 | Generalized Partially Linear Single-Index Models | 522 | 12 |
| y26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y26 | Penalized Spline Estimation For Partially Linear Single-Index Models | 329 | 17 |
| y26 | Optimal Smoothing In Single-Index Models | 386 | 4 |
| y26 | Statistical Estimation In Varying Coefficient Models | 329 | 14 |
| y27 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y27 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y27 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y27 | Doubly Robust Estimation In Missing Data And Causal Inference Models | 422 | 16 |
| y27 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y28 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y28 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y28 | Performance Guarantees For Individualized Treatment Rules | 219 | 22 |
| y28 | Estimating Optimal Treatment Regimes From A Classification Perspective | 117 | 11 |
| y28 | New Statistical Learning Methods For Estimating Optimal Dynamic Treatment Regimes | 96 | 26 |
| y29 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |
| y29 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y29 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y29 | A New Class Of Biased Estimate In Linear-Regression | 257 | 5 |
| y29 | Monte-Carlo Evaluation Of Some Ridge-Type Estimators | 263 | 8 |
| y30 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y30 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y30 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |
| y30 | The Use Of Mixture-Models For The Analysis Of Survival-Data With Long-Term Survivors | 343 | 6 |
| y30 | A Nonparametric Mixture Model For Cure Rate Estimation | 211 | 12 |

Table 7: Z (outgoing citation) factor hubs

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Scad-Penalized Least Absolute Deviation Regression In High-Dimensional Models | 5 | 29 |
| z01 | Regularization Parameter Selections Via Generalized Information Criterion | 102 | 27 |
| z01 | Partially Linear Structure Selection In Cox Models With Varying Coefficients | 11 | 43 |
| z01 | Variable Selection And Estimation Using A Continuous Approximation To The $L$ $_0$ Penalty | 1 | 28 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z02 | $E\ (\chi^2)$-Optimal Mixed-Level Supersaturated Designs | 18 | 32 |
| z02 | Optimal Multi-Level Supersaturated Designs Constructed From Linear And Quadratic Functions | 7 | 23 |

Table 7: Z (outgoing citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z03 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z03 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |
| z03 | Some Recent Work On Multivariate Gaussian Markov Random Fields | 19 | 63 |
| z03 | Bayesian Disease Mapping: Past, Present, And Future | 5 | 98 |
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z04 | Shared Frailty Models For Recurrent Events And A Terminal Event | 173 | 22 |
| z04 | The Adaptive And The Thresholded Lasso For Potentially Misspecified Models (And A Lower Bound For The Lasso) | 42 | 23 |
| z04 | A General Theory Of Concave Regularization For High-Dimensional Sparse Estimation Problems | 83 | 33 |
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | Fifty Years Of The Cox Model | 1 | 43 |
| z05 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z05 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z05 | Nonparametric Inference For Right-Censored Data Using Smoothing Splines | 0 | 37 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z06 | Clustering, Classification, Discriminant Analysis, And Dimension Reduction Via Generalized Hyperbolic Mixtures | 6 | 39 |
| z06 | Mixtures Of Multivariate Contaminated Normal Regression Models | 18 | 63 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Covariate Balancing Propensity Score | 148 | 34 |
| z07 | Balancing Vs Modeling Approaches To Weighting In Practice | 13 | 38 |
| z07 | Globally Efficient Non-Parametric Inference Of Average Treatment Effects By Empirical Balancing Calibration Weighting | 50 | 38 |
| z07 | Calibration Techniques Encompassing Survey Sampling, Missing Data Analysis And Causal Inference | 0 | 74 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z08 | Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error And False Discovery Rates | 20 | 37 |
| z08 | Covariate-Assisted Ranking And Screening For Large-Scale Two-Sample Inference | 29 | 60 |
| z09 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z09 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z09 | Evaluation Of Community-Intervent Ion Trials Via Generalized Linear Mixed Models | 6 | 17 |
| z09 | Monte Carlo Method For Computing The Marginal Likelihood In Nondecomposable Gaussian Graphical Models | 67 | 7 |
| z09 | The Analysis Of Ordered Categorical Data: An Overview And A Survey Of Recent Developments - Discussion | 0 | 174 |
| z10 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z10 | The Nested Dirichlet Process | 69 | 36 |
| z10 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |

Table 7: Z (outgoing citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z10 | Bayesian Dynamic Density Estimation | 33 | 28 |
| z10 | Some Issues In Nonparametric Bayesian Modelling Using Species Sampling Models | 23 | 57 |
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z11 | Covariate Information Matrix For Sufficient Dimension Reduction | 4 | 46 |
| z11 | A Structured Covariance Ensemble For Sufficient Dimension Reduction | 0 | 47 |
| z12 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |
| z12 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z12 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |
| z12 | A Double Ar( $P$ ) Model:: Structure And Estimation | 43 | 30 |
| z12 | Break Detection In The Covariance Structure Of Multivariate Time Series Models | 139 | 27 |
| z13 | A Tutorial On The Lasso Approach To Sparse Modeling | 24 | 16 |
| z13 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z13 | Sparse Partial Least Squares Regression For Simultaneous Dimension Reduction And Variable Selection | 145 | 24 |
| z13 | Penalized Classification Using Fisher's Linear Discriminant | 96 | 24 |
| z13 | Alternative Penalty Functions For Penalized Likelihood Principal Components | 0 | 9 |
| z14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z14 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | Space-Time Covariance Functions | 164 | 17 |
| z14 | A Case Study Competition Among Methods For Analyzing Large Spatial Data | 111 | 52 |
| z14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| z15 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z15 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z15 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z15 | Empirical Likelihood For Single-Index Models | 65 | 29 |
| z15 | Smoothed Empirical Likelihood Methods For Quantile Regression Models | 47 | 30 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z16 | A Modified Pseudolikelihood Approach For Analysis Of Longitudinal Data | 9 | 31 |
| z16 | Effect Partitioning In Cross-Sectionally Clustered Data Without Multilevel Models | 0 | 46 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z17 | Functional Response Models | 70 | 21 |
| z17 | Dynamic Relations For Sparsely Sampled Gaussian Processes | 4 | 59 |
| z17 | Functional Modelling And Classification Of Longitudinal Data | 97 | 45 |
| z18 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z18 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z18 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z18 | On Mixtures Of Skew Normal And Skew T-Distributions | 76 | 28 |
| z18 | Model-Based Clustering And Classification With Non-Normal Mixture Distributions | 38 | 32 |

Table 7: Z (outgoing citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z19 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z19 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z19 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z19 | Model Selection In Binary And Tobit Quantile Regression Using The Gibbs Sampler | 23 | 37 |
| z19 | Linear Quantile Mixed Models | 51 | 45 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Mixtures Of $G$ Priors For Bayesian Variable Selection | 279 | 22 |
| z20 | Transdimensional Markov Chains: A Decade Of Progress And Future Perspectives | 30 | 57 |
| z20 | Hierarchical Bayesian Formulations For Selecting Variables In Regression Models | 9 | 49 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z21 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z21 | Formal And Informal Model Selection With Incomplete Data | 5 | 55 |
| z21 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z22 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z22 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z22 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z22 | Adaptive Designs For Confirmatory Clinical Trials | 90 | 56 |
| z22 | An Adaptive Confirmatory Trial With Interim Treatment Selection: Practical Experiences And Unbalanced Randomization | 6 | 49 |
| z23 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z23 | Twenty Years Of P-Splines | 42 | 103 |
| z23 | Fast Stable Restricted Maximum Likelihood And Marginal Likelihood Estimation Of Semiparametric Generalized Linear Models | 224 | 35 |
| z23 | Generalized Structured Additive Regression Based On Bayesian P-Splines | 116 | 34 |
| z23 | Tutorial In Biostatistics: Spline Smoothing With Linear Mixed Models | 11 | 27 |
| z24 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z24 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z24 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z24 | Large Covariance Estimation By Thresholding Principal Orthogonal Complements | 214 | 63 |
| z24 | An Overview Of The Estimation Of Large Covariance And Precision Matrices | 58 | 58 |
| z25 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |
| z25 | The Hastings Algorithm At Fifty | 7 | 78 |
| z25 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z25 | Riemann Manifold Langevin And Hamiltonian Monte Carlo Methods | 208 | 70 |
| z25 | Unbiased Markov Chain Monte Carlo Methods With Couplings | 24 | 77 |
| z26 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z26 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z26 | Varying Coefficient Regression Models: A Review And New Developments | 40 | 64 |
| z26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| z26 | Efficient Estimation Of A Semiparametric Partially Linear Varying Coefficient Model | 96 | 26 |

Table 7: Z (outgoing citation) factor hubs *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z27 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z27 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z27 | Joint Modeling Of Longitudinal And Survival Data Via A Common Frailty | 22 | 21 |
| z27 | Joint Longitudinal-Survival-Cure Models And Their Application To Prostate Cancer | 64 | 34 |
| z27 | Latent Variable Modelling: A Survey | 20 | 67 |
| z28 | Causal Inference Using Potential Outcomes: Design, Modeling, Decisions | 175 | 32 |
| z28 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z28 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z28 | Defining And Estimating Intervention Effects For Groups That Will Develop An Auxiliary Outcome | 39 | 29 |
| z28 | A Refreshing Account Of Principal Stratification | 18 | 34 |
| z29 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z29 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |
| z29 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z29 | Combining The Unrestricted Estimators Into A Single Estimator And A Simulation Study On The Unrestricted Estimators | 3 | 19 |
| z29 | Consistent Estimation Approach To Tackling Collinearity And Berkson-Type Measurement Error In Linear Regression Using Adjusted Wald-Type Estimator | 2 | 21 |
| z30 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z30 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |
| z30 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |
| z30 | A Flexible Family Of Transformation Cure Rate Models | 13 | 39 |
| z30 | A New Approach To Modeling The Cure Rate In The Presence Of Interval Censored Data | 2 | 45 |

# E Sensitivity to rank

This section contains supplementary results for estimates using varying rank $k$.

## E.1 Y keywords as rank varies

Table 8: Keywords for Y (incoming citation) factors - k = 5, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| y1 | selection, lasso, variable, regression, via, dimensional | y1 |
| y2 | dimension, models, skew, mixture, model, selection | y2 |
| y3 | survival, data, regression, censored, semiparametric, models | y3 |
| y4 | false, discovery, multiple, testing, rate, high | y4 |
| y5 | bayesian, dirichlet, models, spatial, priors, mixture | y5 |

Table 9: Keywords for Y (incoming citation) factors - k = 10, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| y01 | selection, lasso, variable, regression, via, penalized | y01 |
| y02 | bayesian, models, spatial, mixed, longitudinal, monte | y02 |
| y03 | covariance, graphical, sparse, high, dimensional, lasso | y03 |
| y04 | survival, censored, hazards, proportional, data, regression | y04 |
| y05 | mixture, model, clustering, skew, selection, mixtures | y05 |
| y06 | false, discovery, multiple, testing, rate, controlling | y06 |
| y07 | propensity, causal, score, treatment, missing, effects | y07 |
| y08 | dimension, reduction, regression, sliced, inverse, sufficient | y08 |
| y09 | bayesian, dirichlet, nonparametric, mixture, mixtures, priors | y09 |
| y10 | likelihood, empirical, models, regression, longitudinal, quantile | y10 |

Table 10: Keywords for Y (incoming citation) factors - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| y01 | selection, variable, penalized, oracle, nonconcave, lasso | y01 |
| y02 | screening, dimensional, ultrahigh, feature, independence, high | y02 |
| y03 | bayesian, models, complexity, disease, longitudinal, model | y03 |
| y04 | high, dimensional, lasso, selection, models, sparse | y04 |
| y05 | survival, censored, hazards, proportional, cox, data | y05 |
| y06 | model, clustering, mixture, dimension, selection, mixtures | y06 |
| y07 | false, discovery, multiple, testing, rate, controlling | y07 |
| y08 | propensity, causal, treatment, score, observational, effects | y08 |
| y09 | dirichlet, bayesian, nonparametric, mixture, mixtures, priors | y09 |
| y10 | covariance, graphical, sparse, estimation, high, dimensional | y10 |
| y11 | dimension, reduction, regression, sliced, inverse, sufficient | y11 |
| y12 | longitudinal, data, models, generalized, linear, estimating | y12 |
| y13 | spatial, gaussian, models, datasets, covariance, large | y13 |
| y14 | empirical, likelihood, confidence, ratio, semiparametric, partially | y14 |
| y15 | lasso, selection, regression, via, shrinkage, regularization | y15 |
| y16 | functional, regression, longitudinal, linear, principal, data | y16 |
| y17 | skew, normal, distributions, multivariate, distribution, t | y17 |
| y18 | quantile, regression, quantiles, estimation, censored, models | y18 |
| y19 | missing, data, imputation, longitudinal, nonignorable, models | y19 |

| | | |
|---|---|---|
| y20 | bayesian, selection, monte, carlo, variable, bayes | y20 |

Table 11: Keywords for Y (incoming citation) factors - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| y01 | lasso, selection, variable, oracle, elastic, net | y01 |
| y02 | screening, dimensional, ultrahigh, feature, independence, high | y02 |
| y03 | coordinate, descent, regularization, paths, selection, lasso | y03 |
| y04 | bayesian, models, complexity, disease, fit, model | y04 |
| y05 | high, dimensional, lasso, supersaturated, selection, confidence | y05 |
| y06 | survival, censored, hazards, proportional, regression, cox | y06 |
| y07 | model, clustering, mixture, selection, dimension, mixtures | y07 |
| y08 | false, discovery, multiple, rate, testing, controlling | y08 |
| y09 | missing, estimation, semiparametric, treatment, sampling, data | y09 |
| y10 | graphical, covariance, sparse, lasso, estimation, high | y10 |
| y11 | dirichlet, bayesian, nonparametric, mixture, mixtures, priors | y11 |
| y12 | dimension, reduction, regression, sliced, inverse, sufficient | y12 |
| y13 | empirical, likelihood, confidence, ratio, intervals, regions | y13 |
| y14 | propensity, score, causal, effects, observational, treatment | y14 |
| y15 | spatial, covariance, datasets, large, gaussian, temporal | y15 |
| y16 | longitudinal, data, generalized, models, estimating, binary | y16 |
| y17 | functional, regression, principal, data, linear, longitudinal | y17 |
| y18 | lasso, selection, shrinkage, regression, via, variable | y18 |
| y19 | skew, normal, distributions, multivariate, distribution, t | y19 |
| y20 | quantile, regression, quantiles, censored, bayesian, median | y20 |
| y21 | missing, imputation, data, nonignorable, multiple, longitudinal | y21 |
| y22 | bayesian, selection, variable, priors, posterior, prior | y22 |
| y23 | models, spatial, mixed, bayesian, generalized, approximate | y23 |
| y24 | extreme, extremes, tail, max, value, stable | y24 |
| y25 | change, point, detection, series, charts, control | y25 |
| y26 | trials, clinical, adaptive, multiple, sequential, group | y26 |
| y27 | coefficient, models, varying, index, single, partially | y27 |
| y28 | covariance, matrices, high, dimensional, large, matrix | y28 |
| y29 | monte, carlo, markov, metropolis, bayesian, chain | y29 |
| y30 | longitudinal, joint, mixed, models, data, effects | y30 |
| y31 | ridge, regression, biased, estimators, estimator, liu | y31 |
| y32 | garch, series, autoregressive, time, models, volatility | y32 |
| y33 | cure, survival, mixture, rate, model, censored | y33 |
| y34 | treatment, regimes, causal, trials, individualized, learning | y34 |
| y35 | splines, smoothing, additive, regression, penalized, spline | y35 |
| y36 | bayesian, bayes, factors, selection, model, priors | y36 |
| y37 | depth, functional, multivariate, robust, data, outlier | y37 |
| y38 | recurrent, events, event, data, semiparametric, failure | y38 |
| y39 | selection, variable, penalized, nonconcave, oracle, likelihood | y39 |
| y40 | bootstrap, jackknife, nonparametric, lecture, density, estimation | y40 |

Table 12: Keywords for Z (outgoing citation) factors - k = 5, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| z1 | selection, variable, high, dimensional, lasso, regression | z1 |
| z2 | models, model, selection, bayesian, clustering, data | z2 |
| z3 | data, survival, semiparametric, hazards, longitudinal, censored | z3 |
| z4 | false, discovery, high, multiple, testing, dimensional | z4 |
| z5 | bayesian, models, spatial, modeling, dirichlet, hierarchical | z5 |

Table 13: Keywords for Z (outgoing citation) factors - k = 10, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| z01 | selection, variable, high, dimensional, regression, lasso | z01 |
| z02 | bayesian, models, spatial, data, longitudinal, modeling | z02 |
| z03 | high, graphical, dimensional, sparse, covariance, matrix | z03 |
| z04 | survival, hazards, data, censored, cox, model | z04 |
| z05 | selection, model, clustering, mixture, models, skew | z05 |
| z06 | false, discovery, testing, multiple, rate, high | z06 |
| z07 | causal, missing, propensity, treatment, score, effects | z07 |
| z08 | dimension, reduction, sufficient, index, single, regression | z08 |
| z09 | bayesian, dirichlet, nonparametric, mixture, clustering, semiparametric | z09 |
| z10 | empirical, likelihood, quantile, models, longitudinal, data | z10 |

Table 14: Keywords for Z (outgoing citation) factors - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| z01 | selection, variable, dimensional, high, penalized, regression | z01 |
| z02 | screening, adaptive, dimensional, clinical, trials, ultrahigh | z02 |
| z03 | bayesian, models, spatial, model, data, longitudinal | z03 |
| z04 | high, dimensional, selection, lasso, models, sparse | z04 |
| z05 | survival, hazards, censored, data, cox, model | z05 |
| z06 | selection, clustering, model, mixture, models, mixtures | z06 |
| z07 | false, discovery, testing, multiple, rate, microarray | z07 |
| z08 | causal, treatment, propensity, score, effects, observational | z08 |
| z09 | bayesian, dirichlet, nonparametric, mixture, clustering, semiparametric | z09 |
| z10 | graphical, high, covariance, dimensional, sparse, matrix | z10 |
| z11 | dimension, reduction, sufficient, index, single, inverse | z11 |
| z12 | longitudinal, data, models, generalized, mixed, binary | z12 |
| z13 | spatial, models, temporal, spatio, bayesian, gaussian | z13 |
| z14 | empirical, likelihood, inference, partially, missing, models | z14 |
| z15 | selection, lasso, regression, sparse, high, variable | z15 |
| z16 | functional, regression, data, longitudinal, varying, linear | z16 |
| z17 | skew, normal, distributions, multivariate, t, distribution | z17 |
| z18 | quantile, regression, models, varying, index, quantiles | z18 |
| z19 | missing, data, longitudinal, imputation, with, nonignorable | z19 |
| z20 | bayesian, selection, variable, models, priors, model | z20 |

Table 15: Keywords for Z (outgoing citation) factors - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| Factor Name | Top words | ID |
|---|---|---|
| z01 | selection, lasso, variable, high, dimensional, sparse | z01 |
| z02 | screening, dimensional, high, ultrahigh, feature, variable | z02 |
| z03 | adaptive, clinical, trials, sequential, designs, group | z03 |
| z04 | bayesian, models, spatial, model, data, hierarchical | z04 |
| z05 | high, dimensional, supersaturated, designs, selection, regression | z05 |
| z06 | survival, hazards, censored, cox, data, proportional | z06 |
| z07 | selection, clustering, model, mixture, models, mixtures | z07 |
| z08 | false, discovery, testing, multiple, rate, microarray | z08 |
| z09 | missing, treatment, estimation, with, data, robust | z09 |
| z10 | graphical, high, dimensional, sparse, lasso, models | z10 |
| z11 | bayesian, dirichlet, nonparametric, mixture, clustering, semiparametric | z11 |
| z12 | dimension, reduction, sufficient, index, inverse, sliced | z12 |
| z13 | empirical, likelihood, inference, missing, partially, models | z13 |
| z14 | propensity, causal, score, treatment, effects, observational | z14 |
| z15 | spatial, temporal, spatio, gaussian, covariance, modeling | z15 |
| z16 | longitudinal, data, binary, generalized, estimating, clustered | z16 |
| z17 | functional, regression, data, linear, principal, longitudinal | z17 |
| z18 | selection, high, lasso, sparse, dimensional, variable | z18 |
| z19 | skew, normal, distributions, multivariate, t, distribution | z19 |
| z20 | quantile, regression, quantiles, composite, censored, expectile | z20 |
| z21 | missing, imputation, data, longitudinal, with, nonignorable | z21 |
| z22 | bayesian, selection, variable, priors, regression, shrinkage | z22 |
| z23 | mixed, models, spatial, effects, bayesian, data | z23 |
| z24 | extreme, extremes, tail, copula, dependence, value | z24 |
| z25 | change, detection, point, monitoring, high, series | z25 |
| z26 | multiple, trials, clinical, adaptive, testing, procedures | z26 |
| z27 | varying, coefficient, index, models, single, partially | z27 |
| z28 | high, dimensional, covariance, matrices, matrix, factor | z28 |
| z29 | bayesian, monte, carlo, mcmc, metropolis, markov | z29 |
| z30 | longitudinal, joint, mixed, data, models, effects | z30 |
| z31 | ridge, regression, estimator, liu, linear, model | z31 |
| z32 | garch, volatility, models, series, time, autoregressive | z32 |
| z33 | cure, model, survival, rate, data, mixture | z33 |
| z34 | treatment, causal, effects, instrumental, regimes, randomized | z34 |
| z35 | splines, penalized, additive, models, regression, spline | z35 |
| z36 | bayesian, model, selection, models, averaging, bayes | z36 |
| z37 | depth, functional, robust, multivariate, based, data | z37 |
| z38 | recurrent, event, data, events, semiparametric, terminal | z38 |
| z39 | selection, variable, high, dimensional, penalized, regression | z39 |
| z40 | bootstrap, censored, estimation, data, nonparametric, inference | z40 |

## E.2 Z keywords as rank varies

## E.3 Mixing matrix as rank varies

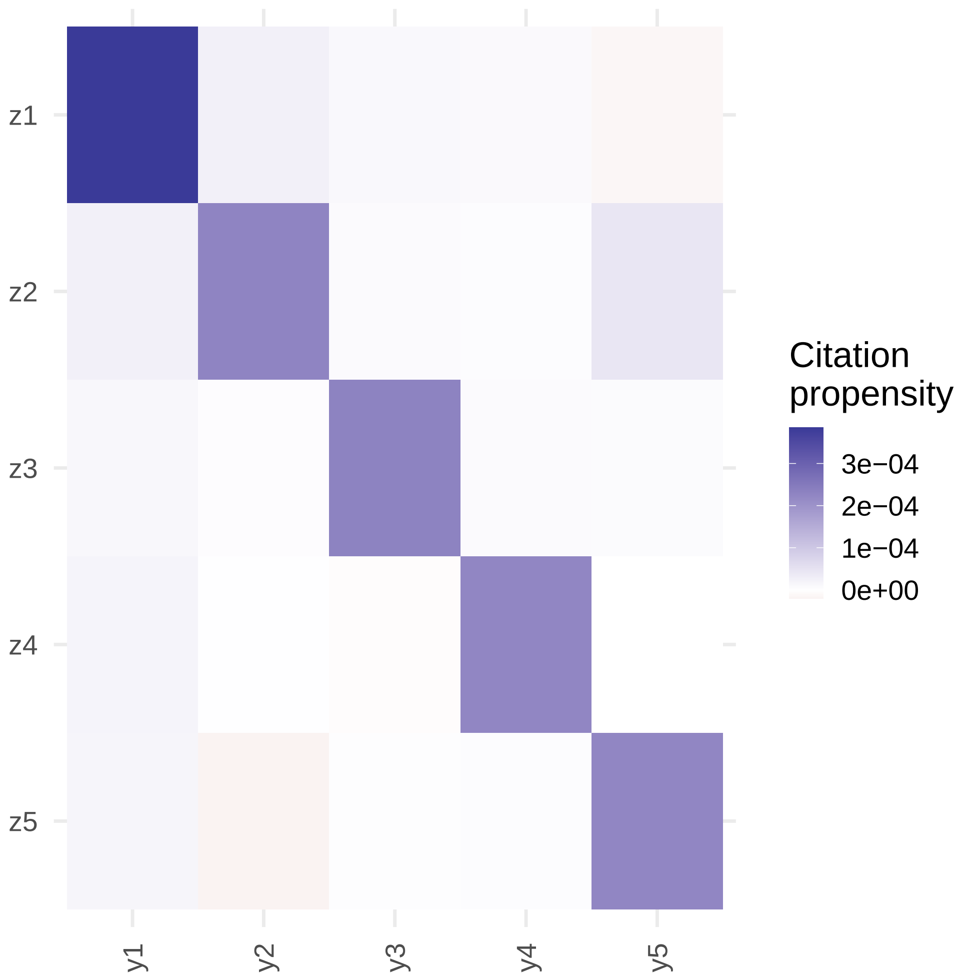


Figure 7: $\widehat{B}$ when $k = 5$

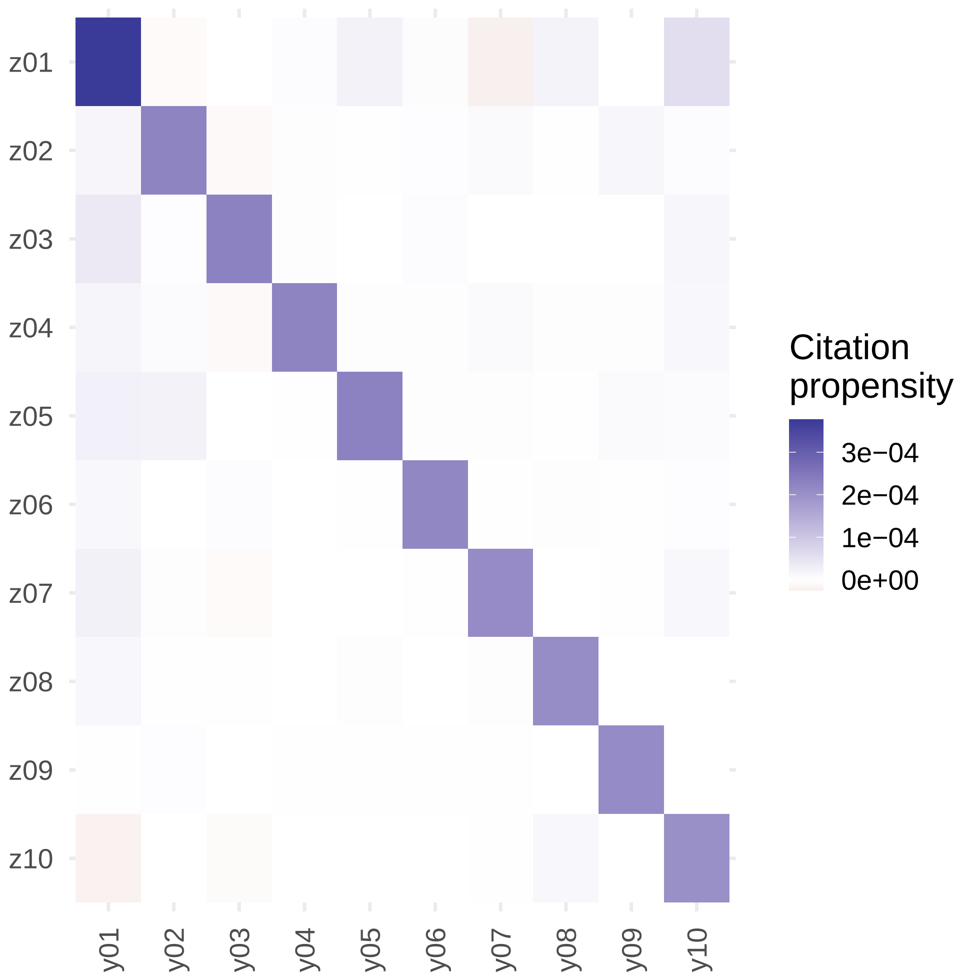


Figure 8: $\widehat{B}$ when $k = 10$

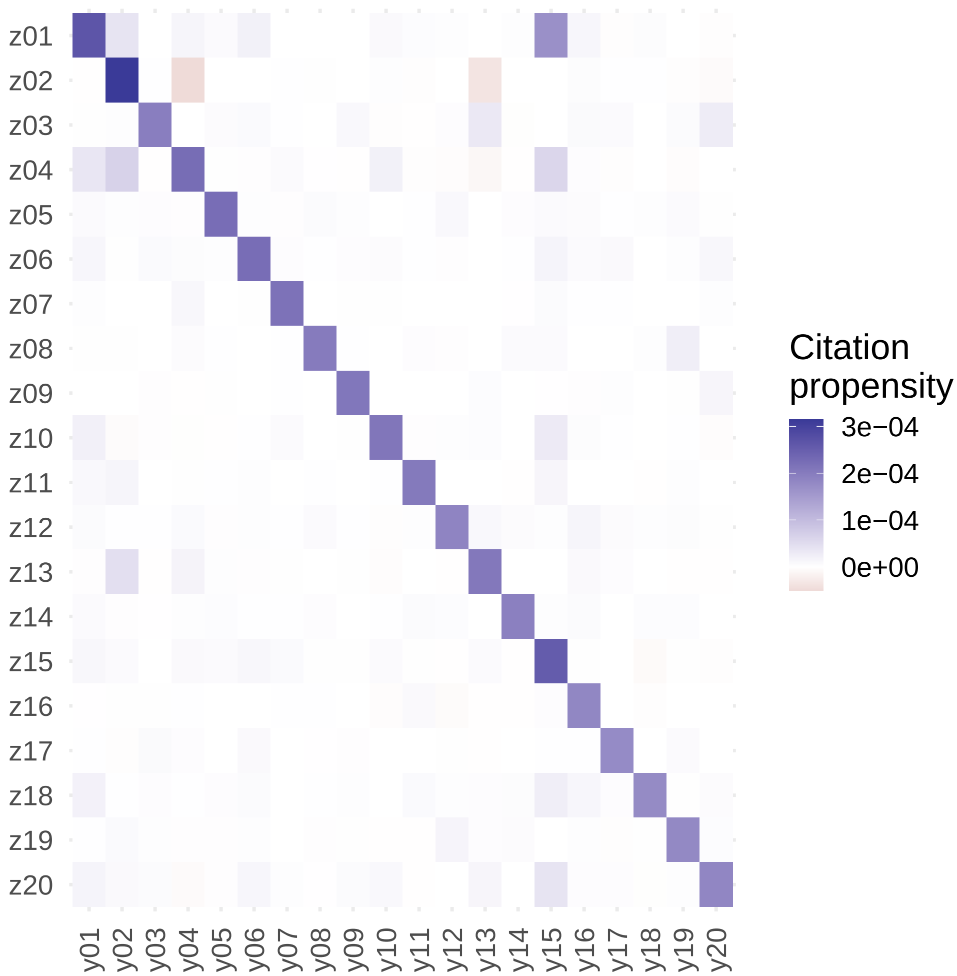


Figure 9: $\widehat{B}$ when $k = 20$

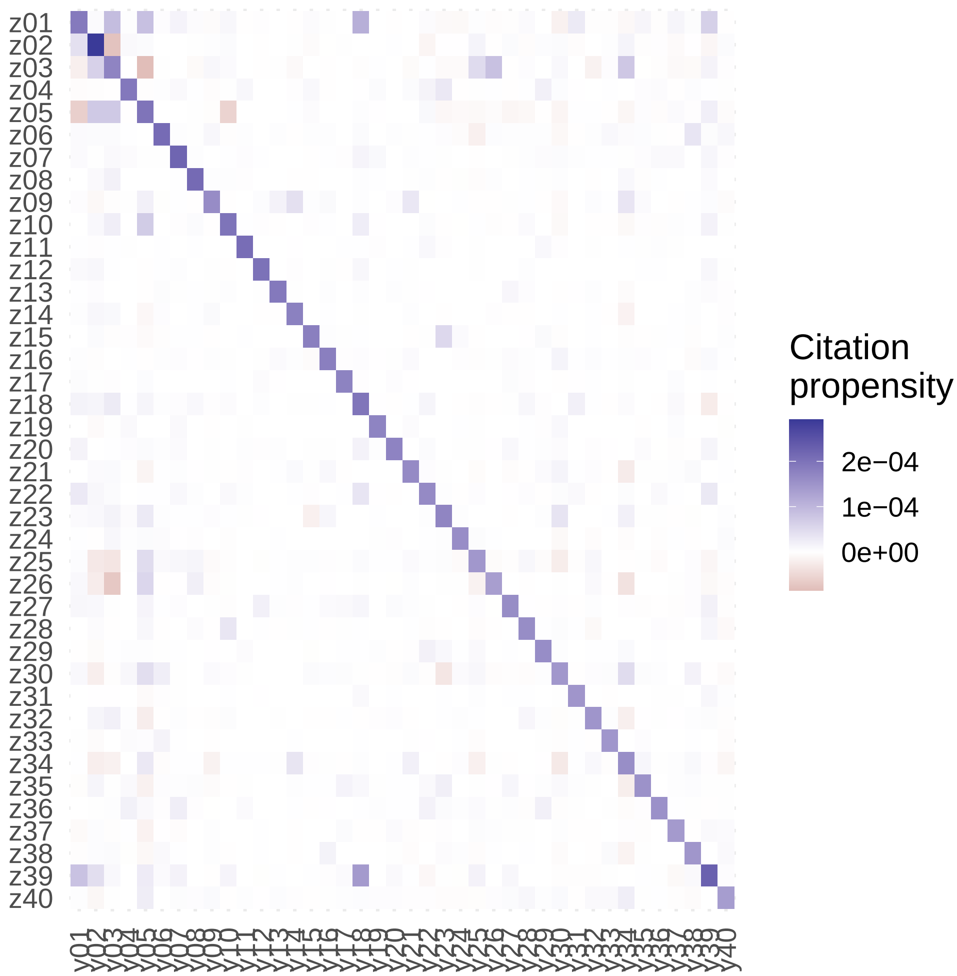


Figure 10: $\widehat{B}$ when $k = 40$

## E.4 Y hubs as rank varies

Table 16: Y (incoming citation) factor hubs - k = 5, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y1 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y1 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y1 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y2 | Estimating Dimension Of A Model | 3727 | 3 |
| y2 | Bayes Factors | 1553 | 56 |
| y2 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y3 | Regression Models And Life-Tables | 4087 | 21 |
| y3 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y3 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y4 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y4 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y4 | A Direct Approach To False Discovery Rates | 555 | 5 |

Table 16: Y (incoming citation) factor hubs - k = 5, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y5 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y5 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y5 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |

Table 17: Y (incoming citation) factor hubs - k = 10, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y02 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y02 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y02 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y03 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y03 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y04 | Regression Models And Life-Tables | 4087 | 21 |
| y04 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y04 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y05 | Estimating Dimension Of A Model | 3727 | 3 |
| y05 | Bayes Factors | 1553 | 56 |
| y05 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y06 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y06 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y07 | Inference And Missing Data | 1905 | 9 |
| y07 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y08 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y08 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y08 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y09 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y09 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y09 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y10 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y10 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y10 | Empirical Likelihood And General Estimating Equations | 877 | 17 |

Table 18: Y (incoming citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |

Table 18: Y (incoming citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y07 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y07 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y07 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y08 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y08 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y08 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y09 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y09 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y09 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y10 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y10 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y10 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y12 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y12 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y12 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y13 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y13 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y13 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y14 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y14 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y14 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y15 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |

Table 18: Y (incoming citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y15 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y15 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y16 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y16 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y16 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y17 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y17 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y17 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y18 | Regression Quantiles | 1603 | 21 |
| y18 | Bayesian Quantile Regression | 286 | 7 |
| y18 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y19 | Inference And Missing Data | 1905 | 9 |
| y19 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y19 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |

Table 19: Y (incoming citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y01 | Least Angle Regression | 1259 | 10 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y03 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y03 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y04 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y04 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y04 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y05 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y05 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y05 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y06 | Regression Models And Life-Tables | 4087 | 21 |
| y06 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Partial Likelihood | 933 | 6 |
| y07 | Estimating Dimension Of A Model | 3727 | 3 |
| y07 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y07 | Some Comments On Cp | 770 | 18 |
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |

Table 19: Y (incoming citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y09 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y09 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y09 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y10 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y10 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y10 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y11 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y11 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y11 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y12 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y12 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y12 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y13 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y13 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y13 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y14 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y14 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| y14 | Reducing Bias In Observational Studies Using Subclassification On The Propensity Score | 313 | 5 |
| y15 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y15 | Fixed Rank Kriging For Very Large Spatial Data Sets | 347 | 15 |
| y15 | Hierarchical Nearest-Neighbor Gaussian Process Models For Large Geostatistical Datasets | 190 | 28 |
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y18 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y18 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y18 | The Bayesian Lasso | 568 | 11 |
| y19 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y19 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y19 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y20 | Regression Quantiles | 1603 | 21 |
| y20 | Bayesian Quantile Regression | 286 | 7 |
| y20 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Mice: Multivariate Imputation By Chained Equations In R | 292 | 34 |
| y21 | Pattern-Mixture Models For Multivariate Incomplete Data | 352 | 14 |
| y22 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y22 | The Bayesian Lasso | 568 | 11 |
| y22 | The Horseshoe Estimator For Sparse Signals | 354 | 14 |
| y23 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |

Table 19: Y (incoming citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y24 | Statistical Modeling Of Spatial Extremes | 188 | 41 |
| y24 | Simple General Approach To Inference About Tail Of A Distribution | 642 | 4 |
| y24 | Statistical-Inference Using Extreme Order Statistics | 598 | 1 |
| y25 | Wild Binary Segmentation For Multiple Change-Point Detection | 172 | 34 |
| y25 | Optimal Detection Of Changepoints With A Linear Computational Cost | 193 | 16 |
| y25 | Continuous Inspection Schemes | 794 | 9 |
| y26 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y26 | A Simple Sequentially Rejective Multiple Test Procedure | 748 | 3 |
| y26 | A Sharper Bonferroni Procedure For Multiple Tests Of Significance | 440 | 4 |
| y27 | Generalized Partially Linear Single-Index Models | 522 | 12 |
| y27 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y27 | Optimal Smoothing In Single-Index Models | 386 | 4 |
| y28 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y28 | Covariance Regularization By Thresholding | 399 | 20 |
| y28 | Determining The Number Of Factors In Approximate Factor Models | 434 | 5 |
| y29 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y29 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y29 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y30 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y30 | A Joint Model For Survival And Longitudinal Data Measured With Error | 409 | 6 |
| y30 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| y31 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |
| y31 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y31 | A New Class Of Biased Estimate In Linear-Regression | 257 | 5 |
| y32 | Autoregressive Conditional Heteroscedasticity With Estimates Of The Variance Of United-Kingdom Inflation | 1299 | 4 |
| y32 | Conditional Heteroskedasticity In Asset Returns - A New Approach | 391 | 9 |
| y32 | Maximum Likelihood Estimation Of Pure Garch And Arma-Garch Processes | 191 | 17 |
| y33 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y33 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y33 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |
| y34 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y34 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y34 | Identification Of Causal Effects Using Instrumental Variables | 659 | 15 |
| y35 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y35 | Bayesian P-Splines | 303 | 27 |
| y35 | Selecting The Number Of Knots For Penalized Splines | 314 | 10 |
| y36 | Bayes Factors | 1553 | 56 |
| y36 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y36 | Bayesian Model Averaging: A Tutorial | 566 | 35 |
| y37 | General Notions Of Statistical Depth Function | 340 | 26 |
| y37 | On A Notion Of Data Depth Based On Random Simplices | 335 | 2 |
| y37 | Multivariate Analysis By Data Depth: Descriptive Statistics, Graphics And Inference | 284 | 34 |

Table 19: Y (incoming citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y38 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y38 | Semiparametric Regression For The Mean And Rate Functions Of Recurrent Events | 386 | 12 |
| y38 | Regression-Analysis Of Multivariate Incomplete Failure Time Data By Modeling Marginal Distributions | 543 | 16 |
| y39 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y39 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y39 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y40 | 1977 Rietz Lecture - Bootstrap Methods - Another Look At The Jackknife | 1796 | 9 |
| y40 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y40 | The Jackknife And The Bootstrap For General Stationary Observations | 520 | 12 |

## E.5 Z hubs as rank varies

Table 20: Z (outgoing citation) factor hubs - k = 5, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z1 | A Survey Of $L_1$ Regression | 10 | 63 |
| z1 | A Selective Overview Of Variable Selection In High Dimensional Feature Space | 293 | 68 |
| z1 | A Majorization-Minimization Approach To Variable Selection Using Spike And Slab Priors | 8 | 36 |
| z2 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z2 | Robust Model-Based Clustering Via Mixtures Of Skew- $T$ Distributions With Missing Information | 7 | 43 |
| z2 | Flexible Mixture Modeling Via The Multivariate T Distribution With The Box-Cox Transformation: An Alternative To The Skew-T Distribution | 24 | 36 |
| z3 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z3 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z3 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z4 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z4 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z4 | Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error And False Discovery Rates | 20 | 37 |
| z5 | Some Issues In Nonparametric Bayesian Modelling Using Species Sampling Models | 23 | 57 |
| z5 | Inference For Mixtures Of Finite Polya Tree Models | 84 | 45 |
| z5 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |

Table 21: Z (outgoing citation) factor hubs - k = 10, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | A Survey Of $L_1$ Regression | 10 | 63 |
| z01 | A Selective Overview Of Variable Selection In High Dimensional Feature Space | 293 | 68 |
| z01 | A Majorization-Minimization Approach To Variable Selection Using Spike And Slab Priors | 8 | 36 |

Table 21: Z (outgoing citation) factor hubs - k = 10, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z02 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Bayesian Disease Mapping: Past, Present, And Future | 5 | 98 |
| z03 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z03 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z03 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z04 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z04 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z05 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z05 | Robust Model-Based Clustering Via Mixtures Of Skew- $T$ Distributions With Missing Information | 7 | 43 |
| z05 | Flexible Mixture Modeling Via The Multivariate T Distribution With The Box-Cox Transformation: An Alternative To The Skew-T Distribution | 24 | 36 |
| z06 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z06 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z06 | Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error And False Discovery Rates | 20 | 37 |
| z07 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| z08 | A Review On Dimension Reduction | 59 | 74 |
| z08 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z08 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z09 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z09 | The Nested Dirichlet Process | 69 | 36 |
| z09 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z10 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z10 | Smoothed Empirical Likelihood Methods For Quantile Regression Models | 47 | 30 |
| z10 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |

Table 22: Z (outgoing citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Sparsistent And Constansistent Estimation Of The Varying-Coefficient Model With A Diverging Number Of Predictors | 3 | 42 |
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Partially Linear Structure Selection In Cox Models With Varying Coefficients | 11 | 43 |
| z02 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z02 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |

Table 22: Z (outgoing citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z03 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z03 | Bayesian Disease Mapping: Past, Present, And Future | 5 | 98 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| z04 | Statistical Significance In High-Dimensional Linear Models | 88 | 33 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z05 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z07 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z07 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z07 | Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error And False Discovery Rates | 20 | 37 |
| z08 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z08 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z08 | Covariate Balancing Propensity Score | 148 | 34 |
| z09 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z09 | The Nested Dirichlet Process | 69 | 36 |
| z09 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z10 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z10 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z10 | Covariance Estimation: The Glm And Regularization Perspectives | 57 | 89 |
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z12 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z12 | Modelling Strategies For Repeated Multiple Response Data | 1 | 41 |
| z12 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z13 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z13 | General Design Bayesian Generalized Linear Mixed Models | 61 | 33 |
| z13 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z14 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z14 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z15 | Boosting Ridge Regression | 23 | 11 |
| z15 | Regularization Approaches In Clinical Biostatistics: A Review Of Methods And Their Applications | 0 | 46 |
| z15 | Usage Of The Go Estimator In High Dimensional Linear Models | 4 | 11 |

Table 22: Z (outgoing citation) factor hubs - k = 20, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z16 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z16 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z16 | Dynamic Relations For Sparsely Sampled Gaussian Processes | 4 | 59 |
| z17 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z17 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z17 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z18 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z18 | Adaptive Varying-Coefficient Linear Quantile Model: A Profiled Estimating Equations Approach | 1 | 43 |
| z18 | Quantile Regression Methods With Varying-Coefficient Models For Censored Data | 9 | 39 |
| z19 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z19 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z19 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Transdimensional Markov Chains: A Decade Of Progress And Future Perspectives | 30 | 57 |

Table 23: Z (outgoing citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | A Survey Of $L_1$ Regression | 10 | 63 |
| z01 | A Selective Overview Of Variable Selection In High Dimensional Feature Space | 293 | 68 |
| z01 | Sparse Regression With Exact Clustering | 36 | 18 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | $E(\chi^2)$-Optimal Mixed-Level Supersaturated Designs | 18 | 32 |
| z02 | Optimal Mixed-Level Supersaturated Designs And A New Class Of Combinatorial Designs | 11 | 27 |
| z03 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z03 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z03 | A 25-Year Review Of Sequential Methodology In Clinical Studies | 11 | 85 |
| z04 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z04 | Some Recent Work On Multivariate Gaussian Markov Random Fields | 19 | 63 |
| z04 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |
| z05 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z05 | Optimal Multi-Level Supersaturated Designs Constructed From Linear And Quadratic Functions | 7 | 23 |
| z05 | Construction Of Some $E(F_{Nod})$ Optimal Mixed-Level Supersaturated Designs | 15 | 22 |
| z06 | *Biometrika* Highlights From Volume 28 Onwards | 2 | 603 |
| z06 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z06 | Nonparametric Inference For Right-Censored Data Using Smoothing Splines | 0 | 37 |
| z07 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |

Table 23: Z (outgoing citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z07 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z07 | Mixtures Of Multivariate Contaminated Normal Regression Models | 18 | 63 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z08 | Power-Enhanced Multiple Decision Functions Controlling Family-Wise Error And False Discovery Rates | 20 | 37 |
| z09 | Calibration Techniques Encompassing Survey Sampling, Missing Data Analysis And Causal Inference | 0 | 74 |
| z09 | Multiply Robust Estimation In Regression Analysis With Missing Data | 75 | 40 |
| z09 | A General Framework For Quantile Estimation With Incomplete Data | 18 | 42 |
| z10 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z10 | Monte Carlo Method For Computing The Marginal Likelihood In Nondecomposable Gaussian Graphical Models | 67 | 7 |
| z10 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z11 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z11 | The Nested Dirichlet Process | 69 | 36 |
| z11 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z12 | A Review On Dimension Reduction | 59 | 74 |
| z12 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z12 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z13 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z13 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z13 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z14 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z14 | An Introduction To Propensity Score Methods For Reducing The Effects Of Confounding In Observational Studies | 36 | 34 |
| z14 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z15 | Space-Time Covariance Functions | 164 | 17 |
| z15 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z15 | A Case Study Competition Among Methods For Analyzing Large Spatial Data | 111 | 52 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | Functional Response Models | 70 | 21 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z18 | Practical Comparison Of Sparse Methods For Classification Of Arabica And Robusta Coffee Species Using Near Infrared Hyperspectral Imaging | 8 | 24 |
| z18 | [Hdda] Sparse Subspace Constrained Partial Least Squares | 0 | 19 |
| z18 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z19 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z19 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z19 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |

Table 23: Z (outgoing citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z20 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z20 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z20 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Bayesian Approaches For Missing Not At Random Outcome Data: The Role Of Identifying Restrictions | 17 | 51 |
| z21 | Formal And Informal Model Selection With Incomplete Data | 5 | 55 |
| z22 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z22 | Hierarchical Bayesian Formulations For Selecting Variables In Regression Models | 9 | 49 |
| z22 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z23 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z23 | Mixed Model Prediction And Small Area Estimation | 127 | 88 |
| z23 | General Design Bayesian Generalized Linear Mixed Models | 61 | 33 |
| z24 | Statistics Of Extremes For Iid Data And Breakthroughs In The Estimation Of The Extreme Value Index: Laurens De Haan Leading Contributions | 16 | 48 |
| z24 | An Overview And Open Research Topics In Statistics Of Univariate Extremes | 34 | 101 |
| z24 | A Modeler's Guide To Extreme Value Software | 0 | 125 |
| z25 | Inference For Single And Multiple Change-Points In Time Series | 47 | 102 |
| z25 | Multiscale Change Point Inference | 118 | 91 |
| z25 | Structural Breaks In Time Series | 137 | 73 |
| z26 | An Adaptive Confirmatory Trial With Interim Treatment Selection: Practical Experiences And Unbalanced Randomization | 6 | 49 |
| z26 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z26 | The Fallback Procedure For Evaluating A Single Family Of Hypotheses | 64 | 16 |
| z27 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z27 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z27 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| z28 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z28 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z28 | Large Covariance Estimation By Thresholding Principal Orthogonal Complements | 214 | 63 |
| z29 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |
| z29 | The Hastings Algorithm At Fifty | 7 | 78 |
| z29 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z30 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z30 | An Approach To Joint Analysis Of Longitudinal Measurements And Competing Risks Failure Time Data | 51 | 31 |
| z30 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z31 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z31 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |
| z31 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z32 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z32 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |

Table 23: Z (outgoing citation) factor hubs - k = 40, $\ell_z$ = 1e+05, $\ell_y$ = 50000 *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z32 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |
| z33 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z33 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |
| z33 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |
| z34 | Defining And Estimating Intervention Effects For Groups That Will Develop An Auxiliary Outcome | 39 | 29 |
| z34 | A Refreshing Account Of Principal Stratification | 18 | 34 |
| z34 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z35 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z35 | Twenty Years Of P-Splines | 42 | 103 |
| z35 | A General Framework For Functional Regression Modelling | 42 | 78 |
| z36 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z36 | Transdimensional Markov Chains: A Decade Of Progress And Future Perspectives | 30 | 57 |
| z36 | Marginal Likelihood Estimation Via Power Posteriors | 88 | 24 |
| z37 | High-Breakdown Robust Multivariate Methods | 85 | 113 |
| z37 | On General Notions Of Depth For Regression | 5 | 62 |
| z37 | A Decomposition Of Total Variation Depth For Understanding Functional Outliers | 5 | 38 |
| z38 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z38 | Semiparametric Transformation Models With Time-Varying Coefficients For Recurrent And Terminal Events | 21 | 29 |
| z38 | Additive-Multiplicative Rates Model For Recurrent Events | 22 | 31 |
| z39 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z39 | Partially Linear Structure Selection In Cox Models With Varying Coefficients | 11 | 43 |
| z39 | Variable Selection In A Partially Linear Proportional Hazards Model With A Diverging Dimensionality | 10 | 31 |
| z40 | Bootstrap Methods For Dependent Data: A Review | 32 | 70 |
| z40 | Stability | 26 | 34 |
| z40 | Peter Hall's Contributions To The Bootstrap | 6 | 43 |

# F Sensitivity to clipping parameters

This section contains supplementary results for estimates using varying clipping factors $\ell_z$ and $\ell_y$, while holding $k = 30$ constant.

## F.1 Y keywords as clipping parameters vary

Table 24: Keywords for Y (incoming citation) factors - k = 30, $\ell_z$ = 1, $\ell_y$ = 1

| Factor Name | Top words | ID |
|---|---|---|
| y01 | series, garch, autoregressive, time, change, extreme | y01 |
| y02 | lasso, regularization, coordinate, longitudinal, descent, selection | y02 |
| y03 | lasso, selection, regression, shrinkage, via, variable | y03 |
| y04 | selection, variable, penalized, oracle, nonconcave, lasso | y04 |
| y05 | graphical, covariance, estimation, sparse, lasso, high | y05 |

| | | |
|---|---|---|
| y06 | functional, regression, data, principal, linear, longitudinal | y06 |
| y07 | recurrent, event, events, longitudinal, semiparametric, joint | y07 |
| y08 | screening, dimensional, ultrahigh, feature, high, independence | y08 |
| y09 | bayesian, bayes, selection, cure, priors, variable | y09 |
| y10 | false, discovery, multiple, testing, rate, controlling | y10 |
| y11 | regression, coefficient, models, varying, estimation, nonparametric | y11 |
| y12 | spatial, gaussian, large, datasets, covariance, spatio | y12 |
| y13 | missing, imputation, data, nonignorable, longitudinal, nonresponse | y13 |
| y14 | quantile, regression, quantiles, censored, estimation, expectile | y14 |
| y15 | covariance, matrices, high, dimensional, matrix, large | y15 |
| y16 | monte, carlo, bayesian, markov, chain, metropolis | y16 |
| y17 | skew, normal, distributions, distribution, multivariate, t | y17 |
| y18 | longitudinal, data, generalized, cluster, binary, estimating | y18 |
| y19 | dimension, reduction, sliced, regression, inverse, sufficient | y19 |
| y20 | longitudinal, depth, joint, data, models, mixed | y20 |
| y21 | bayesian, models, disease, complexity, longitudinal, spatial | y21 |
| y22 | high, dimensional, regression, lasso, depth, dantzig | y22 |
| y23 | empirical, likelihood, confidence, ratio, intervals, jackknife | y23 |
| y24 | models, mixed, generalized, splines, linear, smoothing | y24 |
| y25 | propensity, causal, score, treatment, observational, effects | y25 |
| y26 | model, clustering, selection, mixture, mixtures, dimension | y26 |
| y27 | bayesian, dirichlet, nonparametric, mixture, mixtures, priors | y27 |
| y28 | treatment, regimes, individualized, learning, rules, dynamic | y28 |
| y29 | survival, censored, hazards, proportional, cox, data | y29 |
| y30 | ridge, regression, liu, estimator, biased, nonorthogonal | y30 |

Table 25: Keywords for Y (incoming citation) factors - k = 30, $\ell_z = 25000$, $\ell_y = 25000$

| Factor Name | Top words | ID |
|---|---|---|
| y01 | selection, variable, penalized, oracle, nonconcave, lasso | y01 |
| y02 | screening, dimensional, ultrahigh, feature, independence, high | y02 |
| y03 | graphical, covariance, lasso, sparse, estimation, selection | y03 |
| y04 | high, dimensional, lasso, regression, density, dantzig | y04 |
| y05 | treatment, regimes, learning, individualized, optimal, dynamic | y05 |
| y06 | functional, longitudinal, data, regression, linear, principal | y06 |
| y07 | lasso, selection, regression, shrinkage, via, variable | y07 |
| y08 | bayesian, models, longitudinal, disease, complexity, mapping | y08 |
| y09 | covariance, matrices, high, dimensional, matrix, large | y09 |
| y10 | bayesian, selection, bayes, variable, priors, posterior | y10 |
| y11 | lasso, regularization, censored, coordinate, descent, selection | y11 |
| y12 | models, splines, smoothing, regression, mixed, cure | y12 |
| y13 | models, coefficient, regression, varying, partially, semiparametric | y13 |
| y14 | false, discovery, multiple, testing, rate, controlling | y14 |
| y15 | dimension, reduction, regression, sliced, inverse, sufficient | y15 |
| y16 | survival, censored, hazards, proportional, data, cox | y16 |
| y17 | longitudinal, data, generalized, models, binary, estimating | y17 |
| y18 | bootstrap, cure, nonparametric, censored, jackknife, extreme | y18 |
| y19 | model, clustering, mixture, selection, dimension, mixtures | y19 |
| y20 | trials, clinical, sequential, adaptive, group, designs | y20 |
| y21 | empirical, likelihood, confidence, ratio, missing, intervals | y21 |
| y22 | bayesian, dirichlet, nonparametric, mixture, mixtures, priors | y22 |
| y23 | propensity, causal, score, treatment, observational, effects | y23 |
| y24 | monte, carlo, bayesian, markov, chain, metropolis | y24 |
| y25 | series, change, garch, autoregressive, time, point | y25 |
| y26 | skew, normal, distributions, multivariate, distribution, t | y26 |
| y27 | missing, imputation, data, longitudinal, nonignorable, nonresponse | y27 |

| | | |
|---|---|---|
| y28 | quantile, regression, quantiles, censored, estimation, longitudinal | y28 |
| y29 | spatial, gaussian, datasets, large, spatio, temporal | y29 |
| y30 | ridge, regression, liu, estimator, estimators, biased | y30 |

Table 26: Keywords for Y (incoming citation) factors - k = 30, $\ell_z = 50000$, $\ell_y = 50000$

| Factor Name | Top words | ID |
|---|---|---|
| y01 | selection, variable, penalized, oracle, nonconcave, lasso | y01 |
| y02 | models, spatial, mixed, bayesian, splines, generalized | y02 |
| y03 | graphical, covariance, sparse, estimation, lasso, high | y03 |
| y04 | lasso, selection, regression, shrinkage, via, density | y04 |
| y05 | high, dimensional, lasso, confidence, regression, selection | y05 |
| y06 | treatment, regimes, learning, individualized, optimal, estimating | y06 |
| y07 | bayesian, complexity, models, disease, mapping, fit | y07 |
| y08 | regularization, coordinate, descent, lasso, via, paths | y08 |
| y09 | spatial, large, datasets, gaussian, covariance, bootstrap | y09 |
| y10 | survival, censored, hazards, proportional, regression, cox | y10 |
| y11 | functional, regression, linear, principal, data, longitudinal | y11 |
| y12 | false, discovery, multiple, testing, rate, controlling | y12 |
| y13 | model, clustering, mixture, selection, dimension, mixtures | y13 |
| y14 | longitudinal, data, generalized, models, estimating, binary | y14 |
| y15 | dirichlet, bayesian, nonparametric, mixture, mixtures, priors | y15 |
| y16 | trials, clinical, sequential, adaptive, group, designs | y16 |
| y17 | dimension, reduction, regression, sliced, inverse, sufficient | y17 |
| y18 | screening, dimensional, ultrahigh, feature, independence, high | y18 |
| y19 | propensity, causal, score, effects, treatment, observational | y19 |
| y20 | regression, models, coefficient, varying, nonparametric, semiparametric | y20 |
| y21 | bayesian, selection, bayes, variable, priors, posterior | y21 |
| y22 | empirical, likelihood, confidence, ratio, intervals, regions | y22 |
| y23 | monte, carlo, markov, bayesian, chain, metropolis | y23 |
| y24 | covariance, series, high, matrices, garch, autoregressive | y24 |
| y25 | missing, imputation, data, longitudinal, nonignorable, nonresponse | y25 |
| y26 | quantile, regression, quantiles, censored, estimation, conditional | y26 |
| y27 | skew, normal, distributions, multivariate, distribution, t | y27 |
| y28 | longitudinal, joint, models, mixed, effects, data | y28 |
| y29 | cure, survival, bootstrap, censored, rate, causal | y29 |
| y30 | ridge, regression, biased, estimators, estimator, nonorthogonal | y30 |

Table 27: Keywords for Y (incoming citation) factors - k = 30, $\ell_z = 70000$, $\ell_y = 70000$

| Factor Name | Top words | ID |
|---|---|---|
| y01 | selection, variable, penalized, oracle, nonconcave, lasso | y01 |
| y02 | screening, dimensional, feature, independence, ultrahigh, high | y02 |
| y03 | lasso, high, dimensional, regression, selection, dantzig | y03 |
| y04 | treatment, regimes, censored, survival, individualized, estimating | y04 |
| y05 | spatial, gaussian, bayesian, covariance, datasets, large | y05 |
| y06 | bayesian, complexity, models, disease, fit, mapping | y06 |
| y07 | graphical, covariance, sparse, lasso, estimation, high | y07 |
| y08 | functional, regression, principal, longitudinal, linear, data | y08 |
| y09 | survival, hazards, censored, proportional, regression, cox | y09 |
| y10 | model, dimension, selection, mixture, clustering, estimating | y10 |
| y11 | longitudinal, data, generalized, models, estimating, binary | y11 |
| y12 | dirichlet, bayesian, nonparametric, mixtures, mixture, priors | y12 |
| y13 | covariance, matrices, high, matrix, large, estimation | y13 |

Table 28: Keywords for Z (outgoing citation) factors - $k = 30$, $\ell_z = 1$, $\ell_y = 1$

| Factor Name | Top words | ID |
|---|---|---|
| z01 | skew, density, garch, bibliography, tests, cure | z01 |
| z02 | sequential, trials, clinical, group, adaptive, interim | z02 |
| z03 | mixed, longitudinal, variance, bayesian, effects, components | z03 |
| z04 | selection, high, variable, dimensional, lasso, regression | z04 |
| z05 | wavelet, graphical, covariance, high, bayes, regression | z05 |
| z06 | bootstrap, bootstrapping, confidence, jackknife, functional, resampling | z06 |
| z07 | survival, hazards, censored, proportional, cox, data | z07 |
| z08 | censored, survival, data, kaplan, meier, nonparametric | z08 |
| z09 | bayesian, transformations, transformation, regression, bayes, box | z09 |
| z10 | multiple, false, procedures, discovery, testing, sequential | z10 |
| z11 | density, kernel, estimation, nonparametric, bandwidth, regression | z11 |
| z12 | bayesian, spatial, models, model, survival, disease | z12 |
| z13 | longitudinal, data, missing, covariance, multivariate, growth | z13 |
| z14 | quantile, bootstrap, garch, regression, bayesian, recurrent | z14 |
| z15 | robust, regression, quantile, location, depth, estimators | z15 |
| z16 | bayesian, carlo, monte, sequential, mcmc, quantile | z16 |
| z17 | skew, quantile, regression, propensity, bibliography, normal | z17 |
| z18 | longitudinal, data, causal, contingency, binary, propensity | z18 |
| z19 | reduction, dimension, regression, sufficient, index, single | z19 |
| z20 | designs, longitudinal, models, experiments, experimental, propensity | z20 |
| z21 | unit, root, cointegration, series, autoregressive, roots | z21 |
| z22 | regression, quantile, longitudinal, data, models, nonparametric | z22 |
| z23 | empirical, likelihood, tests, regression, nonparametric, rank | z23 |
| z24 | longitudinal, missing, likelihood, models, data, mixed | z24 |
| z25 | missing, causal, propensity, sampling, imputation, score | z25 |
| z26 | selection, clustering, model, mixture, skew, regression | z26 |
| z27 | bayesian, dirichlet, nonparametric, process, mixture, mixtures | z27 |
| z28 | monte, carlo, bayesian, ridge, metropolis, mcmc | z28 |
| z29 | survival, data, models, hazards, model, censored | z29 |
| z30 | ridge, cure, skew, regression, normal, liu | z30 |

| | | |
|---|---|---|
| y14 | false, discovery, multiple, rate, testing, controlling | y14 |
| y15 | dimension, reduction, regression, sliced, inverse, sufficient | y15 |
| y16 | propensity, causal, score, effects, observational, treatment | y16 |
| y17 | empirical, likelihood, confidence, ratio, intervals, regions | y17 |
| y18 | lasso, selection, regression, shrinkage, via, variable | y18 |
| y19 | coordinate, regularization, descent, lasso, garch, models | y19 |
| y20 | bayesian, selection, variable, bayes, factors, priors | y20 |
| y21 | trials, clinical, sequential, adaptive, group, interim | y21 |
| y22 | quantile, regression, quantiles, censored, median, depth | y22 |
| y23 | missing, data, imputation, longitudinal, nonignorable, drop | y23 |
| y24 | skew, normal, distributions, multivariate, distribution, t | y24 |
| y25 | models, smoothing, regression, longitudinal, mixed, linear | y25 |
| y26 | monte, carlo, markov, chain, metropolis, bayesian | y26 |
| y27 | bootstrap, series, jackknife, cure, autoregressive, models | y27 |
| y28 | models, semiparametric, estimation, coefficient, regression, varying | y28 |
| y29 | cure, survival, mixture, censored, data, proportional | y29 |
| y30 | ridge, regression, biased, estimators, nonorthogonal, estimator | y30 |

Table 29: Keywords for Z (outgoing citation) factors - k = 30, $\ell_z = 25000$, $\ell_y = 25000$

| Factor Name | Top words | ID |
|---|---|---|
| z01 | selection, variable, high, dimensional, lasso, regression | z01 |
| z02 | mixed, models, effects, linear, generalized, data | z02 |
| z03 | graphical, covariance, wavelet, high, skew, dimensional | z03 |
| z04 | density, kernel, bandwidth, estimation, nonparametric, estimators | z04 |
| z05 | unit, root, cointegration, models, roots, tests | z05 |
| z06 | longitudinal, functional, bayesian, data, growth, repeated | z06 |
| z07 | regression, transformation, skew, transformations, selection, data | z07 |
| z08 | bayesian, spatial, models, disease, spatio, hierarchical | z08 |
| z09 | robust, regression, high, breakdown, bayesian, depth | z09 |
| z10 | bayesian, selection, bayes, models, priors, variable | z10 |
| z11 | censored, survival, data, meier, kaplan, censoring | z11 |
| z12 | frailty, bivariate, cure, survival, data, copula | z12 |
| z13 | regression, nonparametric, models, varying, coefficient, estimation | z13 |
| z14 | multiple, false, testing, discovery, procedures, rate | z14 |
| z15 | dimension, reduction, sufficient, index, sliced, single | z15 |
| z16 | survival, hazards, censored, proportional, data, cox | z16 |
| z17 | longitudinal, data, binary, models, estimating, clustered | z17 |
| z18 | bootstrap, bootstrapping, confidence, resampling, intervals, screening | z18 |
| z19 | selection, model, clustering, mixture, models, mixtures | z19 |
| z20 | trials, clinical, sequential, group, adaptive, designs | z20 |
| z21 | empirical, likelihood, inference, semiparametric, models, partially | z21 |
| z22 | bayesian, dirichlet, nonparametric, mixture, process, clustering | z22 |
| z23 | causal, propensity, treatment, score, effects, observational | z23 |
| z24 | bayesian, monte, carlo, markov, chain, gibbs | z24 |
| z25 | garch, series, volatility, time, change, arch | z25 |
| z26 | skew, normal, distributions, multivariate, t, mixtures | z26 |
| z27 | missing, data, longitudinal, imputation, with, incomplete | z27 |
| z28 | quantile, regression, quantiles, estimation, censored, conditional | z28 |
| z29 | recurrent, event, survival, hazards, cox, data | z29 |
| z30 | ridge, regression, estimator, liu, linear, estimators | z30 |

Table 30: Keywords for Z (outgoing citation) factors - k = 30, $\ell_z = 50000$, $\ell_y = 50000$

| Factor Name | Top words | ID |
|---|---|---|
| z01 | selection, variable, lasso, high, dimensional, sparse | z01 |
| z02 | mixed, models, effects, generalized, linear, data | z02 |
| z03 | wavelet, graphical, high, covariance, estimation, dimensional | z03 |
| z04 | density, kernel, bandwidth, regression, nonparametric, estimation | z04 |
| z05 | bayesian, models, gibbs, high, monte, carlo | z05 |
| z06 | unit, root, cointegration, tests, roots, series | z06 |
| z07 | bayesian, spatial, models, disease, mapping, hierarchical | z07 |
| z08 | frailty, survival, recurrent, data, bivariate, copula | z08 |
| z09 | bootstrap, bootstrapping, resampling, confidence, intervals, spatial | z09 |
| z10 | survival, hazards, cox, proportional, data, censored | z10 |
| z11 | robust, functional, regression, depth, high, breakdown | z11 |
| z12 | false, multiple, discovery, testing, rate, microarray | z12 |
| z13 | selection, model, clustering, mixture, models, mixtures | z13 |
| z14 | longitudinal, data, binary, estimating, generalized, clustered | z14 |
| z15 | bayesian, dirichlet, nonparametric, mixture, clustering, semiparametric | z15 |
| z16 | trials, clinical, sequential, group, adaptive, designs | z16 |
| z17 | dimension, reduction, sufficient, index, single, sliced | z17 |
| z18 | censored, survival, data, screening, right, estimation | z18 |
| z19 | causal, propensity, treatment, score, effects, observational | z19 |
| z20 | regression, varying, models, coefficient, nonparametric, partially | z20 |
| z21 | bayesian, selection, bayes, priors, variable, model | z21 |
| z22 | empirical, likelihood, inference, semiparametric, partially, models | z22 |
| z23 | bayesian, monte, carlo, chain, markov, mcmc | z23 |
| z24 | garch, volatility, series, arch, high, time | z24 |
| z25 | missing, data, imputation, longitudinal, with, nonignorable | z25 |
| z26 | quantile, regression, quantiles, censored, conditional, composite | z26 |
| z27 | skew, normal, distributions, multivariate, distribution, t | z27 |
| z28 | longitudinal, mixed, data, effects, models, joint | z28 |
| z29 | cure, survival, data, censored, interval, models | z29 |
| z30 | regression, ridge, estimator, liu, linear, selection | z30 |

Table 31: Keywords for Z (outgoing citation) factors - k = 30, $\ell_z = 70000$, $\ell_y = 70000$

| Factor Name | Top words | ID |
|---|---|---|
| z01 | high, wavelet, dimensional, selection, sparse, regression | z01 |
| z02 | mixed, models, screening, effects, generalized, dimensional | z02 |
| z03 | graphical, high, bayesian, biometrika, spatial, models | z03 |
| z04 | survival, frailty, censored, data, treatment, recurrent | z04 |
| z05 | bayesian, spatial, models, disease, spatio, temporal | z05 |
| z06 | functional, bayesian, regression, high, robust, depth | z06 |
| z07 | longitudinal, mixed, models, data, joint, effects | z07 |
| z08 | functional, bayesian, data, covariance, high, models | z08 |
| z09 | survival, hazards, cox, proportional, data, censored | z09 |
| z10 | selection, model, clustering, mixture, models, mixtures | z10 |
| z11 | longitudinal, data, binary, estimating, generalized, clustered | z11 |
| z12 | bayesian, dirichlet, nonparametric, mixture, clustering, process | z12 |
| z13 | censored, survival, data, estimation, high, covariance | z13 |
| z14 | false, discovery, multiple, testing, rate, microarray | z14 |
| z15 | dimension, reduction, sufficient, index, single, sliced | z15 |
| z16 | propensity, causal, score, treatment, missing, observational | z16 |
| z17 | empirical, likelihood, inference, partially, missing, semiparametric | z17 |
| z18 | selection, variable, lasso, high, dimensional, sparse | z18 |
| z19 | garch, volatility, series, unit, time, root | z19 |
| z20 | bayesian, selection, variable, priors, model, models | z20 |
| z21 | trials, clinical, adaptive, sequential, group, designs | z21 |
| z22 | quantile, regression, quantiles, censored, composite, expectile | z22 |
| z23 | missing, data, imputation, longitudinal, nonignorable, with | z23 |
| z24 | skew, normal, distributions, multivariate, distribution, t | z24 |
| z25 | models, varying, coefficient, regression, partially, functional | z25 |
| z26 | bayesian, monte, carlo, mcmc, markov, metropolis | z26 |
| z27 | bootstrap, recurrent, bootstrapping, models, varying, resampling | z27 |
| z28 | causal, treatment, effects, propensity, instrumental, effect | z28 |
| z29 | cure, survival, data, censored, model, mixture | z29 |
| z30 | regression, ridge, estimator, liu, linear, selection | z30 |

## F.2 Z keywords as clipping parameters vary

## F.3 Mixing matrix as clipping parameters vary

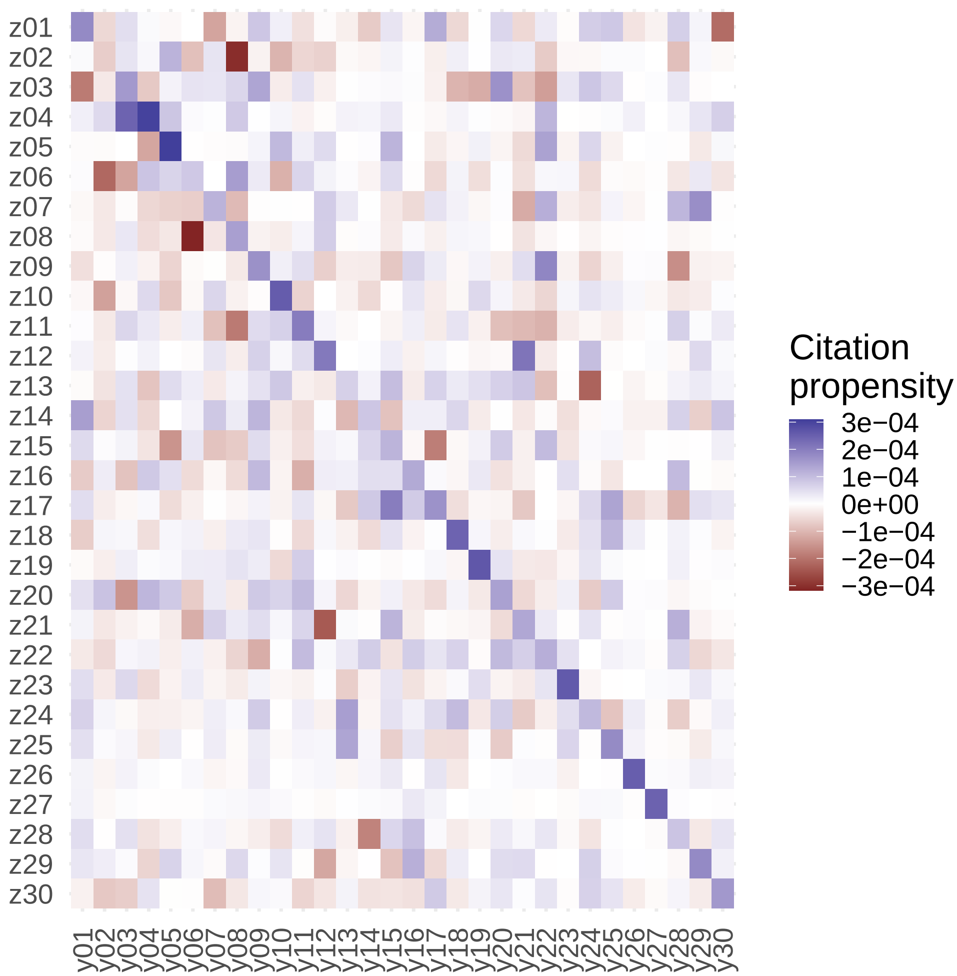


Figure 11: $\widehat{B}$ when $\ell_z = 1$ and $\ell_y = 1$

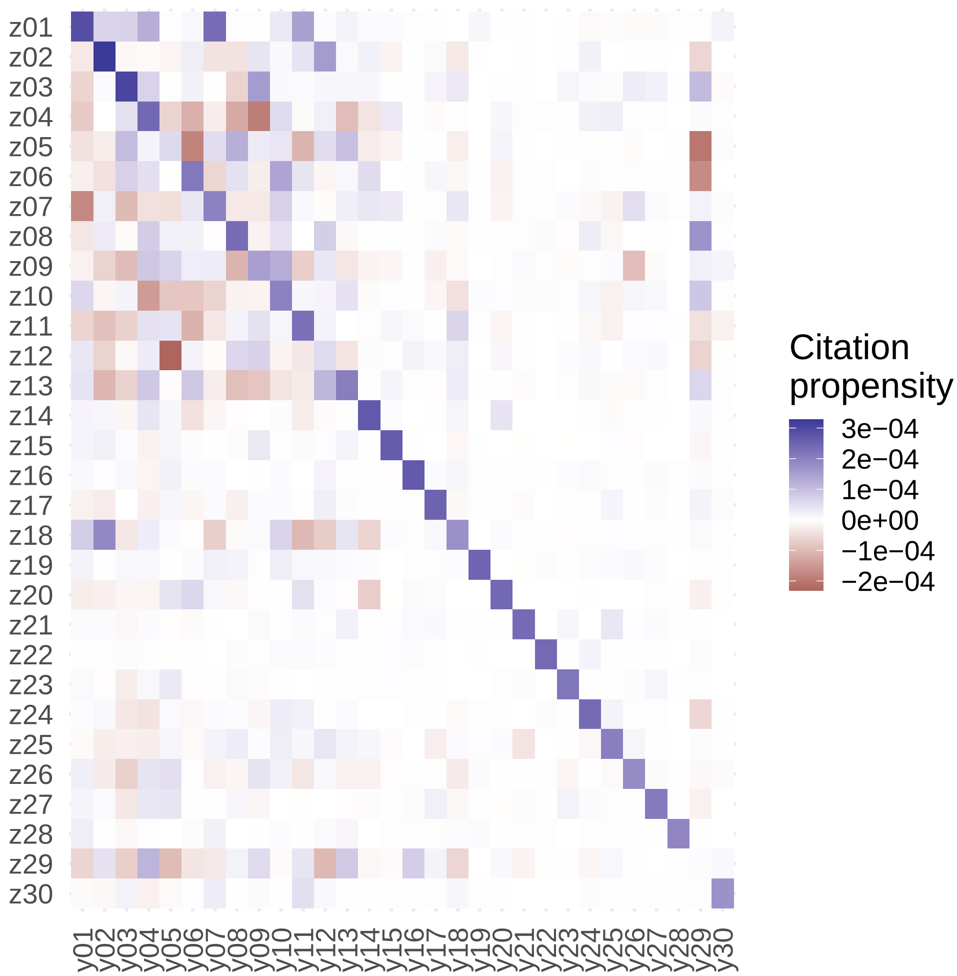


Figure 12: $\widehat{B}$ when $\ell_z = 25,000$ and $\ell_y = 25,000$

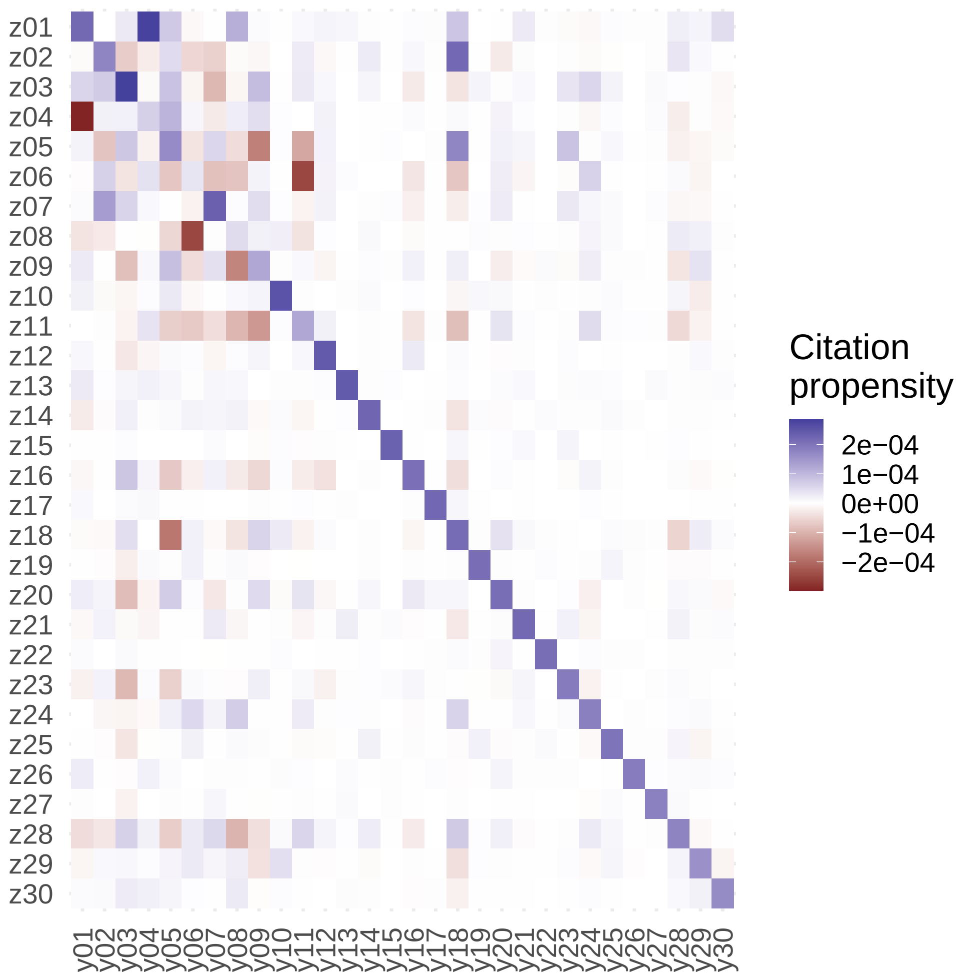


Figure 13: $\widehat{B}$ when $\ell_z = 50,000$ and $\ell_y = 50,000$

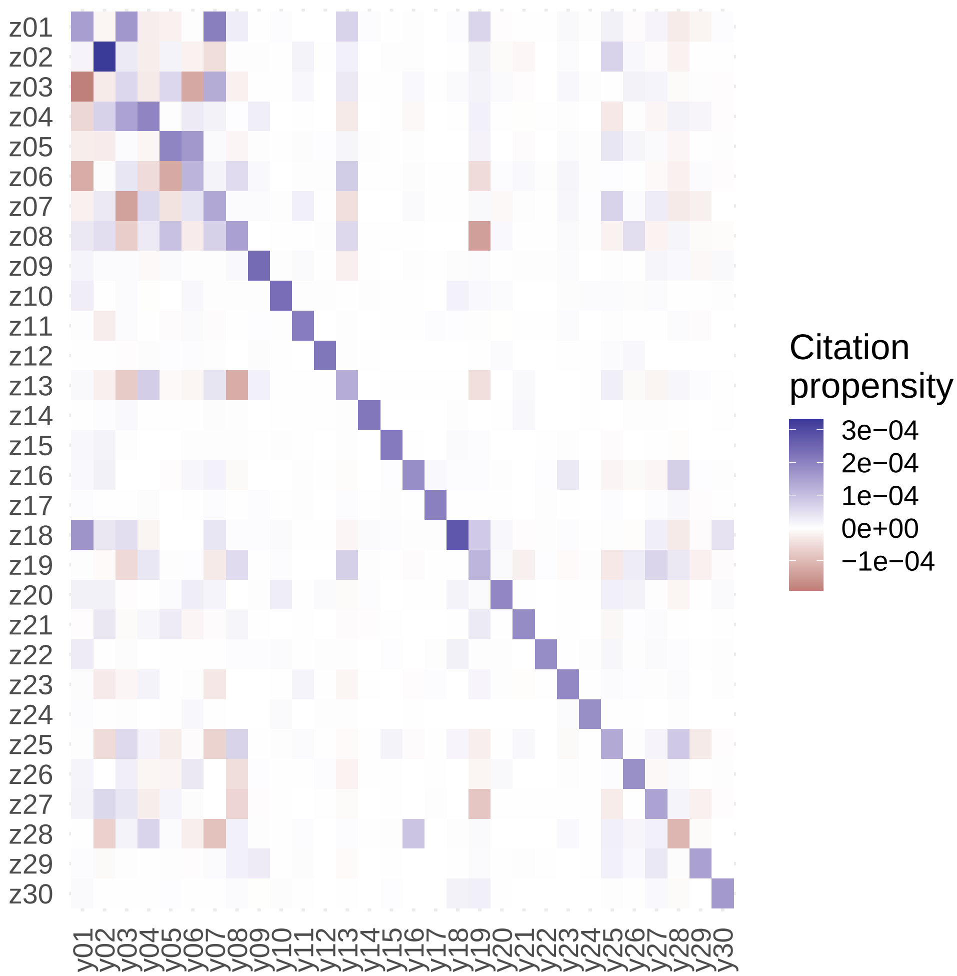


Figure 14: $\widehat{B}$ when $\ell_z = 70,000$ and $\ell_y = 70,000$

## F.4 Y hubs as clipping parameters vary

Table 32: Y (incoming citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |

Table 32: Y (incoming citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y07 | Stratification And Weighting Via The Propensity Score In Estimation Of Causal Treatment Effects: A Comparative Study | 334 | 8 |
| y07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y09 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y09 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y09 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y10 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y10 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y10 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y12 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y12 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y12 | Least Angle Regression | 1259 | 10 |
| y13 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y13 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y13 | Least Angle Regression | 1259 | 10 |
| y14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y14 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y15 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y15 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y15 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y18 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y18 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y18 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |

Table 32: Y (incoming citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y19 | Regression Quantiles | 1603 | 21 |
| y19 | Bayesian Quantile Regression | 286 | 7 |
| y19 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y20 | The Bayesian Lasso | 568 | 11 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y21 | Adjusting For Nonignorable Drop-Out Using Semiparametric Nonresponse Models | 480 | 35 |
| y22 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y22 | Evaluation Of Experiments With Adaptive Interim Analyses | 277 | 7 |
| y22 | Adaptive Sample Size Calculations In Group Sequential Trials | 241 | 16 |
| y23 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y24 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y24 | Covariance Regularization By Thresholding | 399 | 20 |
| y24 | Regularized Estimation Of Large Covariance Matrices | 392 | 14 |
| y25 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y25 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y25 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y26 | Generalized Partially Linear Single-Index Models | 522 | 12 |
| y26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y26 | Penalized Spline Estimation For Partially Linear Single-Index Models | 329 | 17 |
| y27 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y27 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y27 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y28 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y28 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y28 | Performance Guarantees For Individualized Treatment Rules | 219 | 22 |
| y29 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |
| y29 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y29 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y30 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y30 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y30 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |

Table 33: Y (incoming citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y07 | Stratification And Weighting Via The Propensity Score In Estimation Of Causal Treatment Effects: A Comparative Study | 334 | 8 |
| y07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y09 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y09 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y09 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y10 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y10 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y10 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y12 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y12 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y12 | Least Angle Regression | 1259 | 10 |
| y13 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y13 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y13 | Least Angle Regression | 1259 | 10 |
| y14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y14 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |

Table 33: Y (incoming citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y15 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y15 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y15 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y18 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y18 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y18 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y19 | Regression Quantiles | 1603 | 21 |
| y19 | Bayesian Quantile Regression | 286 | 7 |
| y19 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y20 | The Bayesian Lasso | 568 | 11 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y21 | Adjusting For Nonignorable Drop-Out Using Semiparametric Nonresponse Models | 480 | 35 |
| y22 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y22 | Evaluation Of Experiments With Adaptive Interim Analyses | 277 | 7 |
| y22 | Adaptive Sample Size Calculations In Group Sequential Trials | 241 | 16 |
| y23 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y24 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y24 | Covariance Regularization By Thresholding | 399 | 20 |
| y24 | Regularized Estimation Of Large Covariance Matrices | 392 | 14 |
| y25 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y25 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y25 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y26 | Generalized Partially Linear Single-Index Models | 522 | 12 |
| y26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y26 | Penalized Spline Estimation For Partially Linear Single-Index Models | 329 | 17 |
| y27 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y27 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y27 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y28 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y28 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y28 | Performance Guarantees For Individualized Treatment Rules | 219 | 22 |
| y29 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |

Table 33: Y (incoming citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y29 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y29 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y30 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y30 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y30 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |

Table 34: Y (incoming citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y07 | Stratification And Weighting Via The Propensity Score In Estimation Of Causal Treatment Effects: A Comparative Study | 334 | 8 |
| y07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y09 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y09 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y09 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y10 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y10 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y10 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |

Table 34: Y (incoming citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y12 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y12 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y12 | Least Angle Regression | 1259 | 10 |
| y13 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y13 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y13 | Least Angle Regression | 1259 | 10 |
| y14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y14 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y15 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y15 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y15 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y18 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y18 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y18 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y19 | Regression Quantiles | 1603 | 21 |
| y19 | Bayesian Quantile Regression | 286 | 7 |
| y19 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y20 | The Bayesian Lasso | 568 | 11 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y21 | Adjusting For Nonignorable Drop-Out Using Semiparametric Nonresponse Models | 480 | 35 |
| y22 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y22 | Evaluation Of Experiments With Adaptive Interim Analyses | 277 | 7 |
| y22 | Adaptive Sample Size Calculations In Group Sequential Trials | 241 | 16 |
| y23 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y24 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y24 | Covariance Regularization By Thresholding | 399 | 20 |
| y24 | Regularized Estimation Of Large Covariance Matrices | 392 | 14 |
| y25 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y25 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y25 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y26 | Generalized Partially Linear Single-Index Models | 522 | 12 |

Table 34: Y (incoming citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y26 | Penalized Spline Estimation For Partially Linear Single-Index Models | 329 | 17 |
| y27 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y27 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y27 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y28 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y28 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y28 | Performance Guarantees For Individualized Treatment Rules | 219 | 22 |
| y29 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |
| y29 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y29 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y30 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y30 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y30 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |

Table 35: Y (incoming citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y01 | Variable Selection Via Nonconcave Penalized Likelihood And Its Oracle Properties | 2804 | 14 |
| y01 | The Adaptive Lasso And Its Oracle Properties | 2052 | 14 |
| y01 | Nearly Unbiased Variable Selection Under Minimax Concave Penalty | 987 | 36 |
| y02 | Sure Independence Screening For Ultrahigh Dimensional Feature Space | 905 | 33 |
| y02 | Feature Screening Via Distance Correlation Learning | 327 | 22 |
| y02 | Sure Independence Screening In Generalized Linear Models With Np-Dimensionality | 305 | 24 |
| y03 | Bayesian Measures Of Model Complexity And Fit | 2107 | 38 |
| y03 | Prior Distributions For Variance Parameters In Hierarchical Models(Comment On An Article By Browne And Draper) | 906 | 28 |
| y03 | Bayesian Image-Restoration, With 2 Applications In Spatial Statistics | 797 | 3 |
| y04 | On Asymptotically Optimal Confidence Regions And Tests For High-Dimensional Models | 360 | 36 |
| y04 | Confidence Intervals For Low Dimensional Parameters In High Dimensional Linear Models | 350 | 33 |
| y04 | Simultaneous Analysis Of Lasso And Dantzig Selector | 617 | 16 |
| y05 | Regression Models And Life-Tables | 4087 | 21 |
| y05 | Cox Regression-Model For Counting-Processes - A Large Sample Study | 1218 | 13 |
| y05 | Nonparametric-Estimation From Incomplete Observations | 1853 | 11 |
| y06 | Estimating Dimension Of A Model | 3727 | 3 |
| y06 | Model-Based Clustering, Discriminant Analysis, And Density Estimation | 594 | 43 |
| y06 | Some Comments On Cp | 770 | 18 |
| y07 | The Central Role Of The Propensity Score In Observational Studies For Causal Effects | 1497 | 11 |
| y07 | Stratification And Weighting Via The Propensity Score In Estimation Of Causal Treatment Effects: A Comparative Study | 334 | 8 |
| y07 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |

Table 35: Y (incoming citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y08 | Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing | 2025 | 9 |
| y08 | The Control Of The False Discovery Rate In Multiple Testing Under Dependency | 644 | 21 |
| y08 | A Direct Approach To False Discovery Rates | 555 | 5 |
| y09 | Sparse Inverse Covariance Estimation With The Graphical Lasso | 754 | 2 |
| y09 | High-Dimensional Graphs And Variable Selection With The Lasso | 885 | 12 |
| y09 | Model Selection And Estimation In The Gaussian Graphical Model | 477 | 8 |
| y10 | Bayesian Analysis Of Some Nonparametric Problems | 1373 | 4 |
| y10 | A Constructive Definition Of Dirichlet Priors | 711 | 6 |
| y10 | Bayesian Density-Estimation And Inference Using Mixtures | 724 | 12 |
| y11 | Sliced Inverse Regression For Dimension Reduction | 930 | 29 |
| y11 | Sliced Inverse Regression For Dimension Reduction - Comment | 488 | 4 |
| y11 | An Adaptive Estimation Of Dimension Reduction Space | 472 | 25 |
| y12 | Regularization Paths For Generalized Linear Models Via Coordinate Descent | 1124 | 15 |
| y12 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y12 | Least Angle Regression | 1259 | 10 |
| y13 | Regression Shrinkage And Selection Via The Lasso | 4759 | 8 |
| y13 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y13 | Least Angle Regression | 1259 | 10 |
| y14 | An Explicit Link Between Gaussian Fields And Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach | 512 | 55 |
| y14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| y14 | Stationary Process Approximation For The Analysis Of Large Spatial Datasets | 380 | 18 |
| y15 | Empirical Likelihood Ratio Confidence-Intervals For A Single Functional | 1115 | 6 |
| y15 | Empirical Likelihood Ratio Confidence-Regions | 913 | 14 |
| y15 | Empirical Likelihood And General Estimating Equations | 877 | 17 |
| y16 | Longitudinal Data-Analysis Using Generalized Linear-Models | 2624 | 11 |
| y16 | Improving Generalised Estimating Equations Using Quadratic Inference Functions | 235 | 13 |
| y16 | Longitudinal Data-Analysis For Discrete And Continuous Outcomes | 516 | 16 |
| y17 | Functional Data Analysis For Sparse Longitudinal Data | 640 | 24 |
| y17 | Functional Linear Regression Analysis For Longitudinal Data | 342 | 19 |
| y17 | Methodology And Convergence Rates For Functional Linear Regression | 295 | 12 |
| y18 | A Class Of Distributions Which Includes The Normal Ones | 886 | 5 |
| y18 | The Multivariate Skew-Normal Distribution | 632 | 8 |
| y18 | Statistical Applications Of The Multivariate Skew Normal Distribution | 490 | 6 |
| y19 | Regression Quantiles | 1603 | 21 |
| y19 | Bayesian Quantile Regression | 286 | 7 |
| y19 | Composite Quantile Regression And The Oracle Model Selection Theory | 291 | 6 |
| y20 | Bayes Factors | 1553 | 56 |
| y20 | Variable Selection Via Gibbs Sampling | 781 | 8 |
| y20 | The Bayesian Lasso | 568 | 11 |
| y21 | Inference And Missing Data | 1905 | 9 |
| y21 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y21 | Adjusting For Nonignorable Drop-Out Using Semiparametric Nonresponse Models | 480 | 35 |
| y22 | Closed Testing Procedures With Special Reference To Ordered Analysis Of Variance | 557 | 6 |
| y22 | Evaluation Of Experiments With Adaptive Interim Analyses | 277 | 7 |
| y22 | Adaptive Sample Size Calculations In Group Sequential Trials | 241 | 16 |

Table 35: Y (incoming citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| y23 | Flexible Smoothing With B-Splines And Penalties | 1074 | 7 |
| y23 | Approximate Inference In Generalized Linear Mixed Models | 1345 | 41 |
| y23 | Random-Effects Models For Longitudinal Data | 1540 | 11 |
| y24 | On The Distribution Of The Largest Eigenvalue In Principal Components Analysis | 486 | 7 |
| y24 | Covariance Regularization By Thresholding | 399 | 20 |
| y24 | Regularized Estimation Of Large Covariance Matrices | 392 | 14 |
| y25 | Monte-Carlo Sampling Methods Using Markov Chains And Their Applications | 1431 | 4 |
| y25 | Sampling-Based Approaches To Calculating Marginal Densities | 1695 | 9 |
| y25 | Reversible Jump Markov Chain Monte Carlo Computation And Bayesian Model Determination | 1193 | 12 |
| y26 | Generalized Partially Linear Single-Index Models | 522 | 12 |
| y26 | Profile Likelihood Inferences On Semiparametric Varying-Coefficient Partially Linear Models | 360 | 24 |
| y26 | Penalized Spline Estimation For Partially Linear Single-Index Models | 329 | 17 |
| y27 | Estimation Of Regression-Coefficients When Some Regressors Are Not Always Observed | 1034 | 21 |
| y27 | A Generalization Of Sampling Without Replacement From A Finite Universe | 1160 | 5 |
| y27 | Demystifying Double Robustness: A Comparison Of Alternative Strategies For Estimating A Population Mean From Incomplete Data | 416 | 28 |
| y28 | Estimating Individualized Treatment Rules Using Outcome Weighted Learning | 280 | 20 |
| y28 | A Robust Method For Estimating Optimal Treatment Regimes | 210 | 16 |
| y28 | Performance Guarantees For Individualized Treatment Rules | 219 | 22 |
| y29 | Ridge Regression - Biased Estimation For Nonorthogonal Problems | 1402 | 7 |
| y29 | Regularization And Variable Selection Via The Elastic Net | 1584 | 11 |
| y29 | Performance Of Some New Ridge Regression Estimators | 233 | 11 |
| y30 | Estimation In A Cox Proportional Hazards Cure Model | 279 | 10 |
| y30 | Survival Curve For Cancer Patients Following Treatment | 319 | 3 |
| y30 | A Mixture Model Combining Logistic-Regression With Proportional Hazards Regression | 258 | 5 |

## F.5 Z hubs as clipping parameters vary

Table 36: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Scad-Penalized Least Absolute Deviation Regression In High-Dimensional Models | 5 | 29 |
| z01 | Regularization Parameter Selections Via Generalized Information Criterion | 102 | 27 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z03 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z03 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |

Table 36: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z04 | Shared Frailty Models For Recurrent Events And A Terminal Event | 173 | 22 |
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | Fifty Years Of The Cox Model | 1 | 43 |
| z05 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Covariate Balancing Propensity Score | 148 | 34 |
| z07 | Balancing Vs Modeling Approaches To Weighting In Practice | 13 | 38 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z09 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z09 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z09 | Evaluation Of Community-Intervent Ion Trials Via Generalized Linear Mixed Models | 6 | 17 |
| z10 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z10 | The Nested Dirichlet Process | 69 | 36 |
| z10 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z12 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |
| z12 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z12 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |
| z13 | A Tutorial On The Lasso Approach To Sparse Modeling | 24 | 16 |
| z13 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z13 | Sparse Partial Least Squares Regression For Simultaneous Dimension Reduction And Variable Selection | 145 | 24 |
| z14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z14 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | Space-Time Covariance Functions | 164 | 17 |
| z15 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z15 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z15 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |

Table 36: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z17 | Functional Response Models | 70 | 21 |
| z18 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z18 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z18 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z19 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z19 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z19 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Mixtures Of *G* Priors For Bayesian Variable Selection | 279 | 22 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z21 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z22 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z22 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z22 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z23 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z23 | Twenty Years Of P-Splines | 42 | 103 |
| z23 | Fast Stable Restricted Maximum Likelihood And Marginal Likelihood Estimation Of Semiparametric Generalized Linear Models | 224 | 35 |
| z24 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z24 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z24 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z25 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |
| z25 | The Hastings Algorithm At Fifty | 7 | 78 |
| z25 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z26 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z26 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z26 | Varying Coefficient Regression Models: A Review And New Developments | 40 | 64 |
| z27 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z27 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z27 | Joint Modeling Of Longitudinal And Survival Data Via A Common Frailty | 22 | 21 |
| z28 | Causal Inference Using Potential Outcomes: Design, Modeling, Decisions | 175 | 32 |
| z28 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z28 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z29 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z29 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |

Table 36: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 1$, $\ell_y = 1$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z29 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z30 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z30 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |
| z30 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |

Table 37: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Scad-Penalized Least Absolute Deviation Regression In High-Dimensional Models | 5 | 29 |
| z01 | Regularization Parameter Selections Via Generalized Information Criterion | 102 | 27 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z03 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z03 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z04 | Shared Frailty Models For Recurrent Events And A Terminal Event | 173 | 22 |
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | Fifty Years Of The Cox Model | 1 | 43 |
| z05 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Covariate Balancing Propensity Score | 148 | 34 |
| z07 | Balancing Vs Modeling Approaches To Weighting In Practice | 13 | 38 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z09 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z09 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z09 | Evaluation Of Community-Intervent Ion Trials Via Generalized Linear Mixed Models | 6 | 17 |
| z10 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z10 | The Nested Dirichlet Process | 69 | 36 |
| z10 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |

Table 37: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z12 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |
| z12 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z12 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |
| z13 | A Tutorial On The Lasso Approach To Sparse Modeling | 24 | 16 |
| z13 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z13 | Sparse Partial Least Squares Regression For Simultaneous Dimension Reduction And Variable Selection | 145 | 24 |
| z14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z14 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | Space-Time Covariance Functions | 164 | 17 |
| z15 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z15 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z15 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z17 | Functional Response Models | 70 | 21 |
| z18 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z18 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z18 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z19 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z19 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z19 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Mixtures Of *G* Priors For Bayesian Variable Selection | 279 | 22 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z21 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z22 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z22 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z22 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z23 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z23 | Twenty Years Of P-Splines | 42 | 103 |

Table 37: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 25000$, $\ell_y = 25000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z23 | Fast Stable Restricted Maximum Likelihood And Marginal Likelihood Estimation Of Semiparametric Generalized Linear Models | 224 | 35 |
| z24 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z24 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z24 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z25 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |
| z25 | The Hastings Algorithm At Fifty | 7 | 78 |
| z25 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z26 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z26 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z26 | Varying Coefficient Regression Models: A Review And New Developments | 40 | 64 |
| z27 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z27 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z27 | Joint Modeling Of Longitudinal And Survival Data Via A Common Frailty | 22 | 21 |
| z28 | Causal Inference Using Potential Outcomes: Design, Modeling, Decisions | 175 | 32 |
| z28 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z28 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z29 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z29 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |
| z29 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z30 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z30 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |
| z30 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |

Table 38: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Scad-Penalized Least Absolute Deviation Regression In High-Dimensional Models | 5 | 29 |
| z01 | Regularization Parameter Selections Via Generalized Information Criterion | 102 | 27 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z03 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z03 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z04 | Shared Frailty Models For Recurrent Events And A Terminal Event | 173 | 22 |

Table 38: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | Fifty Years Of The Cox Model | 1 | 43 |
| z05 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Covariate Balancing Propensity Score | 148 | 34 |
| z07 | Balancing Vs Modeling Approaches To Weighting In Practice | 13 | 38 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z09 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z09 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z09 | Evaluation Of Community-Intervent Ion Trials Via Generalized Linear Mixed Models | 6 | 17 |
| z10 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z10 | The Nested Dirichlet Process | 69 | 36 |
| z10 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z12 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |
| z12 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z12 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |
| z13 | A Tutorial On The Lasso Approach To Sparse Modeling | 24 | 16 |
| z13 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z13 | Sparse Partial Least Squares Regression For Simultaneous Dimension Reduction And Variable Selection | 145 | 24 |
| z14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z14 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | Space-Time Covariance Functions | 164 | 17 |
| z15 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z15 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z15 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z17 | Functional Response Models | 70 | 21 |

Table 38: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 50000$, $\ell_y = 50000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z18 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z18 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z18 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z19 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z19 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z19 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Mixtures Of *G* Priors For Bayesian Variable Selection | 279 | 22 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z21 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z22 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z22 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z22 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z23 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z23 | Twenty Years Of P-Splines | 42 | 103 |
| z23 | Fast Stable Restricted Maximum Likelihood And Marginal Likelihood Estimation Of Semiparametric Generalized Linear Models | 224 | 35 |
| z24 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z24 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z24 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z25 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |
| z25 | The Hastings Algorithm At Fifty | 7 | 78 |
| z25 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z26 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z26 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z26 | Varying Coefficient Regression Models: A Review And New Developments | 40 | 64 |
| z27 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z27 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z27 | Joint Modeling Of Longitudinal And Survival Data Via A Common Frailty | 22 | 21 |
| z28 | Causal Inference Using Potential Outcomes: Design, Modeling, Decisions | 175 | 32 |
| z28 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z28 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z29 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z29 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |
| z29 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z30 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z30 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |

| | | | |
|---|---|---|---|
| z30 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |

Table 39: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z01 | Scad-Penalized Regression In Additive Partially Linear Proportional Hazards Models With An Ultra-High-Dimensional Linear Part | 3 | 37 |
| z01 | Scad-Penalized Least Absolute Deviation Regression In High-Dimensional Models | 5 | 29 |
| z01 | Regularization Parameter Selections Via Generalized Information Criterion | 102 | 27 |
| z02 | Bayesian Statistics In Medicine: A 25 Year Review | 23 | 511 |
| z02 | Construction Of Optimal Multi-Level Supersaturated Designs | 47 | 32 |
| z02 | Methodological Issues With Adaptation Of Clinical Trial Design | 12 | 41 |
| z03 | Are Nonprofit Antipoverty Organizations Located Where They Are Needed? A Spatial Analysis Of The Greater Hartford Region | 1 | 21 |
| z03 | Space-Time Interactions In Bayesian Disease Mapping With Recent Tools: Making Things Easier For Practitioners | 0 | 25 |
| z03 | Bayesian Shared Spatial-Component Models To Combine And Borrow Strength Across Sparse Disease Surveillance Sources | 2 | 25 |
| z04 | Maximum Likelihood Estimation In Semiparametric Regression Models With Censored Data | 204 | 53 |
| z04 | Estimation For High-Dimensional Linear Mixed-Effects Models Using $L_1$-Penalization | 47 | 24 |
| z04 | Shared Frailty Models For Recurrent Events And A Terminal Event | 173 | 22 |
| z05 | An Overview Of Semiparametric Models In Survival Analysis | 5 | 77 |
| z05 | Fifty Years Of The Cox Model | 1 | 43 |
| z05 | Marginal Screening For High-Dimensional Predictors Of Survival Outcomes | 2 | 56 |
| z06 | Asymmetric Clusters And Outliers: Mixtures Of Multivariate Contaminated Shifted Asymmetric Laplace Distributions | 11 | 66 |
| z06 | The Multivariate Leptokurtic-Normal Distribution And Its Application In Model-Based Clustering | 25 | 56 |
| z06 | A Mixture Of Generalized Hyperbolic Distributions | 60 | 39 |
| z07 | Matching Methods For Causal Inference: A Review And A Look Forward | 272 | 66 |
| z07 | Covariate Balancing Propensity Score | 148 | 34 |
| z07 | Balancing Vs Modeling Approaches To Weighting In Practice | 13 | 38 |
| z08 | A Review Of Modern Multiple Hypothesis Testing, With Particular Attention To The False Discovery Proportion | 40 | 70 |
| z08 | Multiple Hypothesis Testing In Genomics | 21 | 65 |
| z08 | Adaptive False Discovery Rate Control For Heterogeneous Data | 7 | 37 |
| z09 | Experiments In Stochastic Computation For High-Dimensional Graphical Models | 95 | 12 |
| z09 | Covariance Matrix Selection And Estimation Via Penalised Normal Likelihood | 159 | 19 |
| z09 | Evaluation Of Community-Intervent Ion Trials Via Generalized Linear Mixed Models | 6 | 17 |
| z10 | Mixture Models With A Prior On The Number Of Components | 70 | 67 |
| z10 | The Nested Dirichlet Process | 69 | 36 |
| z10 | A Comparative Review Of Variable Selection Techniques For Covariate Dependent Dirichlet Process Mixture Models | 9 | 43 |
| z11 | A Review On Dimension Reduction | 59 | 74 |
| z11 | Sufficient Dimension Reduction Via Inverse Regression: A Minimum Discrepancy Approach | 179 | 35 |
| z11 | Feature Filter For Estimating Central Mean Subspace And Its Sparse Solution | 1 | 53 |
| z12 | Modeling Multiple Regimes In Financial Volatility With A Flexible Coefficient Garch(1,1) Model | 9 | 30 |
| z12 | Automated Inference And Learning In Modeling Financial Volatility | 34 | 24 |
| z12 | Pseudo-Maximum Likelihood Estimation Of Arch($\infty$) Models | 27 | 22 |

Table 39: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z13 | A Tutorial On The Lasso Approach To Sparse Modeling | 24 | 16 |
| z13 | Regularized Partial Least Squares With An Application To Nmr Spectroscopy | 6 | 22 |
| z13 | Sparse Partial Least Squares Regression For Simultaneous Dimension Reduction And Variable Selection | 145 | 24 |
| z14 | Approximate Bayesian Inference For Latent Gaussian Models By Using Integrated Nested Laplace Approximations | 812 | 109 |
| z14 | High-Dimensional Bayesian Geostatistics | 23 | 51 |
| z14 | Space-Time Covariance Functions | 164 | 17 |
| z15 | A Review Of Empirical Likelihood Methods For Time Series | 18 | 76 |
| z15 | A Review On Empirical Likelihood Methods For Regression | 81 | 63 |
| z15 | Empirical Likelihood For A Varying Coefficient Model With Longitudinal Data | 133 | 34 |
| z16 | Improving The Correlation Structure Selection Approach For Generalized Estimating Equations And Balanced Longitudinal Data | 19 | 29 |
| z16 | Finite Sample Adjustments In Estimating Equations And Covariance Estimators For Intracluster Correlations | 21 | 32 |
| z16 | Criterion For The Simultaneous Selection Of A Working Correlation Structure And Either Generalized Estimating Equations Or The Quadratic Inference Function Approach | 10 | 34 |
| z17 | Methods For Scalar-On-Function Regression | 58 | 124 |
| z17 | A Survey Of Functional Principal Component Analysis | 24 | 88 |
| z17 | Functional Response Models | 70 | 21 |
| z18 | An Overview On The Progeny Of The Skew-Normal Family-A Personal Perspective | 4 | 104 |
| z18 | A Unified View On Skewed Distributions Arising From Selections | 106 | 37 |
| z18 | The Skew-Normal Distribution And Related Multivariate Families | 255 | 36 |
| z19 | Posterior Inference In Bayesian Quantile Regression With Asymmetric Laplace Likelihood | 33 | 49 |
| z19 | Multiple Quantile Modeling Via Reduced-Rank Regression | 1 | 37 |
| z19 | Bayesian Model Selection In Ordinal Quantile Regression | 6 | 49 |
| z20 | Bayesian Approaches To Variable Selection: A Comparative Study From Practical Perspectives | 4 | 78 |
| z20 | Prior Distributions For Objective Bayesian Analysis | 27 | 141 |
| z20 | Mixtures Of *G* Priors For Bayesian Variable Selection | 279 | 22 |
| z21 | A Brief Review Of Approaches To Non-Ignorable Non-Response | 4 | 68 |
| z21 | Missing-Data Methods For Generalized Linear Models: A Comparative Review | 148 | 64 |
| z21 | Analysis Of Longitudinal Data With Drop-Out: Objectives, Assumptions And A Proposal | 40 | 53 |
| z22 | Group Sequential And Adaptive Designs - A Review Of Basic Concepts And Points Of Discussion | 12 | 76 |
| z22 | Twenty-Five Years Of Confirmatory Adaptive Designs: Opportunities And Pitfalls | 53 | 113 |
| z22 | Adaptive Seamless Designs: Selection And Prospective Testing Of Hypotheses | 31 | 61 |
| z23 | Semiparametric Regression During 2003-2007 | 58 | 219 |
| z23 | Twenty Years Of P-Splines | 42 | 103 |
| z23 | Fast Stable Restricted Maximum Likelihood And Marginal Likelihood Estimation Of Semiparametric Generalized Linear Models | 224 | 35 |
| z24 | Random Matrix Theory In Statistics: A Review | 53 | 148 |
| z24 | Estimating Structured High-Dimensional Covariance And Precision Matrices: Optimal Rates And Adaptive Estimation | 63 | 96 |
| z24 | Recent Developments In High Dimensional Covariance Estimation And Its Related Issues, A Review | 2 | 54 |
| z25 | Bayesian Computation: A Summary Of The Current State, And Samples Backwards And Forwards | 20 | 101 |

Table 39: Z (outgoing citation) factor hubs - k = 30, $\ell_z = 70000$, $\ell_y = 70000$ *(continued)*

| ID | Title | Cited by | Cites |
|---|---|---|---|
| z25 | The Hastings Algorithm At Fifty | 7 | 78 |
| z25 | A Short History Of Markov Chain Monte Carlo: Subjective Recollections From Incomplete Data | 17 | 68 |
| z26 | Nonparametric Inference With Generalized Likelihood Ratio Tests | 46 | 67 |
| z26 | Statistical Inference In Partially-Varying-Coefficient Single-Index Model | 26 | 27 |
| z26 | Varying Coefficient Regression Models: A Review And New Developments | 40 | 64 |
| z27 | Joint Modeling Of Longitudinal And Time-To-Event Data: An Overview | 341 | 36 |
| z27 | Missing Data Methods In Longitudinal Studies: A Review | 76 | 79 |
| z27 | Joint Modeling Of Longitudinal And Survival Data Via A Common Frailty | 22 | 21 |
| z28 | Causal Inference Using Potential Outcomes: Design, Modeling, Decisions | 175 | 32 |
| z28 | Instrumental Variables: An Econometrician's Perspective | 25 | 88 |
| z28 | Causal Inference: A Missing Data Perspective | 21 | 104 |
| z29 | Modified Liu-Type Estimator Based On (Rk) Class Estimator | 17 | 18 |
| z29 | A Simulation Study On Some Restricted Ridge Regression Estimators | 12 | 17 |
| z29 | Performance Of Kibria's Method For The Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence | 20 | 24 |
| z30 | Piecewise Linear Approximations For Cure Rate Models And Associated Inferential Issues | 20 | 44 |
| z30 | Proportional Hazards Under Conway-Maxwell-Poisson Cure Rate Model And Associated Inference | 8 | 31 |
| z30 | A Support Vector Machine-Based Cure Rate Model For Interval Censored Data | 0 | 50 |