



Biblioteca del Instituto de Estadística y Cartografía de Andalucía

Resúmenes de revistas
Mayo 2015



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO

PRESENTACIÓN

El presente boletín de resúmenes tiene una periodicidad mensual y con él la Biblioteca del Instituto de Estadística y Cartografía de Andalucía pretende dar a conocer a los usuarios de una forma detallada el contenido de las revistas especializadas que entran en su colección. Se trata de un complemento al boletín de novedades de publicaciones seriadas ya que en él se incluyen los resúmenes de cada uno de los artículos que aparecen publicados en los diferentes números de las revistas en el idioma original de las mismas.

Los resúmenes de este boletín corresponden a las revistas que han ingresado en la Biblioteca del Instituto de Estadística y Cartografía de Andalucía durante el mes de **mayo de 2015** y que pueden consultarse gratuitamente en sus instalaciones en la siguiente dirección:

Instituto de Estadística y Cartografía de Andalucía
Pabellón de Nueva Zelanda
C/Leonardo Da Vinci, n. 21. Isla de La Cartuja
41071 - SEVILLA
E-mail: biblio.ieca@juntadeandalucia.es
Teléfono: 955 033 800
Fax: 955 033 816

Horario de atención al público:

Lunes y martes: de 9:00h a 14:00h. y de 16:00 a 19:00 h.

Miércoles, jueves y viernes: de 9:00h a 14:00h.

Horario de verano (del 15 de junio al 15 de septiembre), Semana Santa, Feria de Sevilla y Navidad (del 24 de diciembre al 6 de enero): de lunes a viernes de 9:00h. a 14:00h.



Annals of the Institute of Statistical Mathematics, ISSN 0020-3157
Volume 67, number 3 (june 2015)

Testing regression models with selection-biased data

P. 411-436

J. L. Ojeda - W. González-Manteiga

Abstract

In this paper, we study integrated regression techniques to check the adequacy of a given model in the context of selection-biased observations. We introduce integrated regression in this setting, providing not only a suitable statistic for enabling a model checking test, but also a bootstrap distributional approximation to carry out the test. We also address the behaviour of the test under different alternatives showing that this behaviour is asymptotically the same for both selection-biased and non selection-biased data. The technique is illustrated with a simulation study and a data analysis based on a real situation that shows the performance of the method and how selection bias affect both estimation and inference.

Testing for additivity in nonparametric quantile regression

P. 437-477

Holger Dette - Matthias Gühlich

Abstract

In this article, we propose a new test for additivity in nonparametric quantile regression with a high-dimensional predictor. Asymptotic normality of the corresponding test statistic (after appropriate standardization) is established under the null hypothesis, local and fixed alternatives. We also propose a bootstrap procedure which can be used to improve the approximation of the nominal level for moderate sample sizes. The methodology is also illustrated by means of a small simulation study, and a data example is analyzed.

Robust conditional Weibull-type estimation

P. 479-514

Yuri Goegebeur - Armelle Guillou

Abstract

We study nonparametric robust tail coefficient estimation when the variable of interest, assumed to be of Weibull type, is observed simultaneously with a random covariate. In particular, we introduce a robust estimator for the tail coefficient, using the idea of the density power divergence, based on the relative excesses above a high threshold. The main asymptotic properties of our estimator are established under very general assumptions. The finite sample performance of the proposed procedure is evaluated by a small simulation experiment.

Spacings around an order statistic

P. 515-540

H. N. Nagaraja - Karthik Bharath

Abstract

We determine the joint limiting distribution of adjacent spacings around a central, intermediate, or an extreme order statistic T_n of a random sample of size n from a continuous distribution F . For central and intermediate cases, normalized spacings in the left and right neighborhoods are asymptotically i.i.d. exponential random variables. The

associated independent Poisson arrival processes are independent of TeX . For an extreme TeX , the asymptotic independence property of spacings fails for TeX in the domain of attraction of Fréchet and Weibull (TeX) distributions. This work also provides additional insight into the limiting distribution for the number of observations around TeX for all three cases.

On the equivariance criterion in statistical prediction

P. 541-555

Haojin Zhou - Tapan K. Nayak

Abstract

This paper presents a general development of the basic logic of equivariance for a parametric point prediction problem. We propose a framework that allows the set of possible predictions as well as the losses to depend on the data and then explore the nature and properties of relevant transformation groups for applying the functional and formal equivariance principles. We define loss invariance and predictive equivariance appropriately and discuss their ramifications. We describe a structure of equivariant predictors in terms of maximal invariants and present a method for deriving minimum risk equivariant predictors. We explore the connections between equivariance and risk unbiasedness and show that uniquely best risk unbiased predictors are almost equivariant. We apply our theoretical results to some illustrative examples.

Probabilistic properties of second order branching process

P. 557-572

Akanksha S. Kashikar - S. R. Deshmukh

Abstract

The classical BGW process assumes first order dependence, whereas many real life datasets exhibit a second or higher order dependence. Further, in some situations, there is a need for a model which allows for simultaneous reproduction by a parent and its offspring. This paper proposes a second order branching process model to accommodate such situations and discusses its probabilistic properties such as extinction probability and limiting behaviour of the generation sizes. Estimation of offspring means and growth rate are also discussed. This model is further used to model the swine flu data for Pune, India, and La-Gloria, Mexico.

The sinh-arcsinh logistic family of distributions: properties and inference

P. 573-594

Arthur Pewsey - Toshihiro Abe

Abstract

The sinh-arcsinh transform is used to obtain a flexible four-parameter model that provides a natural framework with which to perform inference robust to wide-ranging departures from the logistic distribution. Its basic properties are established and its distribution and quantile functions, and properties related to them, shown to be highly tractable. Two important subfamilies are also explored. Maximum likelihood estimation is discussed, and reparametrisations designed to reduce the asymptotic correlations between the maximum likelihood estimates provided. A likelihood-ratio test for logisticness, which outperforms standard empirical distribution function based tests, follows naturally. The application of the proposed model and inferential methods is illustrated in an analysis of carbon fibre strength data. Multivariate extensions of the model are explored.

Smooth change point estimation in regression models with random design

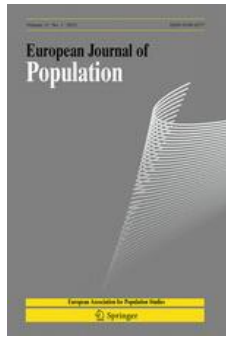
P. 595-619

Maik Döring - Uwe Jensen

Abstract

We consider the problem of estimating the location of a change point θ_0 in a regression model. Most change point models studied so far were based on regression functions with a jump. However, we focus on regression functions, which are continuous at θ_0 . The degree of smoothness q_0 has to be estimated as well. We investigate the

consistency with increasing sample size n of the least squares estimates $(\hat{\theta}_n, \hat{\varphi}_n)$ of (θ_0, φ_0) . It turns out that the rates of convergence of $\hat{\theta}_n$ depend on φ_0 : for φ_0 greater than $1/2$ we have a rate of $n^{-\varphi_0}$ and the asymptotic normality property; for φ_0 less than $1/2$ the rate is $n^{-1/(2\varphi_0+1)}$ and the change point estimator converges to a maximizer of a Gaussian process; for φ_0 equal to $1/2$ the rate is $n \cdot \ln(n)^{-1/2}$. Interestingly, in the last case the limiting distribution is also normal



European journal of population, ISSN 0168-6577
Volume 31, number 2 (may 2015)

**Special issue on Demographic issues of intra-European migration:
destinations, family and settlement**

EDITORIAL NOTES

P. 109-125

Demographic Issues of Intra-European Migration: Destinations, Family and Settlement

Teresa Castro-Martín, Clara Cortina

Labor Market Laws and Intra-European Migration: The Role of the State in Shaping Destination Choices

P127-153

John R. B. Palmer - Mariola Pytliková

Abstract

This article investigates the relationship between migrants' destination choices and the formal labor market access afforded by multiple potential host countries in the context of the EU's eastward enlargement. We use an index of labor market access laws combined with data on migration from new EU member states into the existing states of the EU and EFTA from 2004 through 2010 to test whether (1) migrants are attracted to destinations that give them greater formal labor market access, (2) migration flows to any given destination are influenced by the labor market policies of competing destinations, and (3) the influence of labor market laws on migrant flows is mediated by social networks, language ability, and educational level. Our data support the first two propositions and partly support the third: Migration between origin/destination pairs was positively associated with the loosening of destination labor market restrictions, while negatively associated with the loosening of competing destinations' labor market restrictions. In addition, the influence of destination labor market access appears to be weaker for destinations in which migrants have larger existing co-national networks, and for migrants from countries with languages that are more similar to the destination language, although we do not discern a clear mediating effect of education level. Our models also include variables for a set of economic indicators, social welfare spending, geographic distance, and historical relationships, and the estimated coefficients on these variables are largely in line with theoretical predictions. By combining rich EU data with a unique approach to evaluating competing legal regimes, the analysis helps us better understand how law shapes migration in a multi-destination world.

Life Paths of Migrants: A Sequence Analysis of Polish Migrants' Family Life Trajectories

P. 155-179

Tom Kleinepijper - Helga A. G. de Valk - Ruben van Gaalen

Abstract

After Poland's accession to the European Union in 2004, Polish migration within Europe has increased considerably. Still, little is known about this large and growing group of immigrants. Using unique panel data from the Dutch population registers, we first apply sequence analysis to explore and describe the variety of family life courses among young adult (aged 20–30 at migration) Polish migrants in the Netherlands. Each individual family life career is followed for a 5-year period on a monthly time base. Optimal matching analysis is applied to compute (dis)similarities between individuals, in order to cluster into predominant life paths. Second, a series of logistic regression analyses were carried out to determine the main factors of importance on the different family life paths. Findings showed a strong impact of the partner's origin on the type of union

and likelihood of return migration. Unemployment was found to be another important trigger of return migration. With regard to gender differences, we found that men are more likely to remain single and to leave the Netherlands at some point, while women more often cohabit and tend to stay.

Beyond l'Auberge Espagnole: The Effect of Individual Mobility on the Formation of Intra-European Couples

P. 181-206

Julia H. Schroedter - Tom De Winter - Suzana Koelet

Abstract

This paper focuses on intra-European partnership formation in three European countries: Belgium, the Netherlands and Switzerland. Intra-European mobility has been actively promoted and stimulated by the European Commission (e.g., free movement of persons, the Erasmus student exchange program). One of the reasons for this promotion is that exchanges and relationships between Europeans of different descent are seen as a core indicator of the success of the European project. In this paper, we address the question to what extent intra-European mobility fosters partnerships between Europeans of different descent. Intra-European mobility can create opportunities both to meet partners from other European countries and to accumulate the necessary capital (economic, cultural, linguistic, mobility) to engage in a relationship with a foreign European. We use original data on European (binational) couples, collected in 2012 in the three countries (EUMARR survey), to study the choice of native men and women to engage in a relationship with either a foreign-born European partner or a partner from the own native country. We apply a broader life course perspective that captures migration and mobility experiences prior to the relationship as causal antecedents leading to an intra-European partnership. Results based on logistic regression models suggest that there is an individual effect of long stays abroad and short mobility experiences in (early) adulthood on having an intra-European partner (in comparison with a native partner).

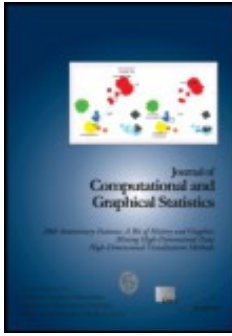
Between Flows and Places: Using Geodemographics to Explore EU Migration Across Neighbourhoods in Britain

P. 207-230

Albert Sabater

Abstract

During the past decade, the growing significance of EU migrants in Britain has attracted considerable media and policy attention. While the rhetoric and public policy debate has focused on national and regional levels, surprisingly, little is known about the emerging settlement patterns of EU migrants at the local level. This paper sheds some light on this important issue by exploring the socio-spatial characteristics of the places of destination/residence of EU migrants and, in doing so, reveals the extent to which the concentration-dispersal framework applies to their settlement pattern. To evaluate whether and in what ways EU migrants have settled across neighbourhoods in Britain, 2001 and 2011 census estimates as well as geodemographic data are used. The findings suggest that EU migrants have consolidated their national presence in Britain because of their growing numbers and unprecedented geographical dispersal. While EU migrants' settlement does not seem to translate into strong clustering patterns, diverging socio-spatial experiences are found among the largest groups (Polish, Italian, Portuguese and Spanish).



Journal of computational and graphical statistics, ISSN 1061-8600
Volume 24, number 1 (march 2015)

ARTICLES

Advances in Dendrogram Seriation for Application to Visualization

P. 1-25

Denise Earle - Catherine B. Hurley

Abstract

Visualizations of statistical data benefit from systematic ordering of data objects to highlight features and structure. This article concerns ordering via dendrogram seriation based on hierarchical clustering of data objects. It describes DendSer, a general-purpose dendrogram seriation algorithm which when coupled with various seriation cost functions is easily adapted to different visualization settings. Comparisons are made with other dendrogram seriation algorithms and applications are presented. Supplementary materials for this article are available online.

Visualizing Complex Data With Embedded Plots

P. 26-43

Garrett Grolemond - Hadley Wickham

Abstract

This article describes a class of graphs, embedded plots, that are particularly useful for analyzing large and complex datasets. Embedded plots organize a collection of graphs into a larger graphic, which can display more complex relationships than would otherwise be possible. This arrangement provides additional axes, prevents overplotting, and allows for multiple levels of visual summarization. Embedded plots also preprocess complex data into a form suitable for the human cognitive system, which can facilitate comprehension. We illustrate the usefulness of embedded plots with a case study, discuss the practical and cognitive advantages of embedded plots, and demonstrate how to implement embedded plots as a general class within visualization software, something currently unavailable. This article has supplementary material online.

Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation

P. 44-65

Alex Goldstein - Adam Kapelner - Justin Bleich - Emil Pitkin

Abstract

This article presents individual conditional expectation (ICE) plots, a tool for visualizing the model estimated by any supervised learning algorithm. Classical partial dependence plots (PDPs) help visualize the average partial relationship between the predicted response and one or more features. In the presence of substantial interaction effects, the partial response relationship can be heterogeneous. Thus, an average curve, such as the PDP, can

obfuscate the complexity of the modeled relationship. Accordingly, ICE plots refine the PDP by graphing the functional relationship between the predicted response and the feature for *individual* observations. Specifically, ICE plots highlight the variation in the fitted values across the range of a covariate, suggesting where and to what extent heterogeneities might exist. In addition to providing a plotting suite for exploratory analysis, we include a visual test for additive structure in the data-generating model. Through simulated examples and real datasets, we demonstrate how ICE plots can shed light on estimated models in ways PDPs cannot. Procedures outlined are available in the R package ICEbox.

Covariance-Guided Mixture Probabilistic Principal Component Analysis (C-MPPCA)

P. 66-83

Chao Han - Scotland Leman - Leanna House

Abstract

To extract information from high-dimensional data efficiently, visualization tools based on data projection methods have been developed and shown useful. However, a single two-dimensional visualization is often insufficient for capturing all or most interesting structures in complex high-dimensional datasets. For this reason, Tipping and Bishop developed mixture probabilistic principal component analysis (MPPCA) that separates data into multiple groups and enables a unique projection per group; that is, one probabilistic principal component analysis (PPCA) data visualization per group. Because the group labels are assigned to observations based on their high-dimensional coordinates, MPPCA works well to reveal homoscedastic structures in data that differ spatially. In the presence of heteroscedasticity, however, MPPCA may still mask noteworthy data structures. We propose a new method called covariance-guided MPPCA (C-MPPCA) that groups subsets of observations based on covariance, not locality, and, similar to MPPCA, displays them using PPCA. PPCA projects data in the dimensions with the highest variances, thus grouping by covariance makes sense and enables some data structures to be visible that were masked originally by MPPCA. We demonstrate the performance of C-MPPCA in an extensive simulation study. We also apply C-MPPCA to a real world dataset. Supplementary materials for this article are available online.

Integrating Data Transformation in Principal Components Analysis

P. 84-103

Mehdi Maadooliat - Jianhua Z. Huang - Jianhua Hu

Abstract

Principal component analysis (PCA) is a popular dimension-reduction method to reduce the complexity and obtain the informative aspects of high-dimensional datasets. When the data distribution is skewed, data transformation is commonly used prior to applying PCA. Such transformation is usually obtained from previous studies, prior knowledge, or trial-and-error. In this work, we develop a model-based method that integrates data transformation in PCA and finds an appropriate data transformation using the maximum profile likelihood. Extensions of the method to handle functional data and missing values are also developed. Several numerical algorithms are provided for efficient computation. The proposed method is illustrated using simulated and real-world data examples. Supplementary materials for this article are available online.

Stable Estimation in Dimension Reduction

P. 104-120

Wenbo Wu - Xiangrong Yin

Abstract

We introduce stable estimation procedures for several aspects of a sufficient dimension-reduction matrix. We first propose a stable method for estimating structural dimension, which only selects the correct directions in the central

subspace with no false positive selection. We then provide a Grassmann manifold sparse estimate for the central subspace. By using subsampling, we develop an ensemble method to obtain a stable nonsparse estimate for the central subspace. This ensemble idea is also used to stabilize the choice of the number of slices in sliced inverse methods. Theoretical results are established, and the efficacy of the proposed stable methods is demonstrated by simulation studies and the analysis of Hitters' salary data. Supplementary materials for this article are available online.

**High-Dimensional Fused Lasso Regression Using Majorization–
Minimization and Parallel Processing**

P. 121-153

Donghyeon Yu - Joong-Ho Won - Taehoon Lee - Johan Lim - Sungroh Yoon

Abstract

In this article, we propose a majorization–minimization (MM) algorithm for high-dimensional fused lasso regression (FLR) suitable for parallelization using graphics processing units (GPUs). The MM algorithm is stable and flexible as it can solve the FLR problems with various types of design matrices and penalty structures within a few tens of iterations. We also show that the convergence of the proposed algorithm is guaranteed. We conduct numerical studies to compare our algorithm with other existing algorithms, demonstrating that the proposed MM algorithm is competitive in many settings including the two-dimensional FLR with arbitrary design matrices. The merit of GPU parallelization is also exhibited. Supplementary materials are available online.

A Multiobjective Exploratory Procedure for Regression Model Selection

P. 154-182

Ankur Sinha - Pekka Malo - Timo Kuosmanen

Abstract

Variable selection is recognized as one of the most critical steps in statistical modeling. The problems encountered in engineering and social sciences are commonly characterized by over-abundance of explanatory variables, nonlinearities, and unknown interdependencies between the regressors. An added difficulty is that the analysts may have little or no prior knowledge on the relative importance of the variables. To provide a robust method for model selection, this article introduces the multiobjective genetic algorithm for variable selection (MOGA-VS) that provides the user with an optimal set of regression models for a given dataset. The algorithm considers the regression problem as a two objective task, and explores the Pareto-optimal (best subset) models by preferring those models over the other which have less number of regression coefficients and better goodness of fit. The model exploration can be performed based on in-sample or generalization error minimization. The model selection is proposed to be performed in two steps. First, we generate the frontier of Pareto-optimal regression models by eliminating the dominated models without any user intervention. Second, a decision-making process is executed which allows the user to choose the most preferred model using visualizations and simple metrics. The method has been evaluated on a recently published real dataset on Communities and Crime Within the United States.

Graphical Models for Ordinal Data

P. 183-204

Jian Guo - Elizaveta Levina - George Michailidis - Ji Zhu

Abstract

This article considers a graphical model for ordinal variables, where it is assumed that the data are generated by discretizing the marginal distributions of a latent multivariate Gaussian distribution. The

relationships between these ordinal variables are then described by the underlying Gaussian graphical model and can be inferred by estimating the corresponding concentration matrix. Direct estimation of the model is computationally expensive, but an approximate EM-like algorithm is developed to provide an accurate estimate of the parameters at a fraction of the computational cost. Numerical evidence based on simulation studies shows the strong performance of the algorithm, which is also illustrated on datasets on movie ratings and an educational survey.

A Localized Implementation of the Iterative Proportional Scaling Procedure for Gaussian Graphical Models

P. 205-229

Ping-Feng Xu - Jianhua Guo - Man-Lai Tang

Abstract

In this article, we propose localized implementations of the iterative proportional scaling (IPS) procedure by the strategy of partitioning cliques for computing maximum likelihood estimations in large Gaussian graphical models. We first divide the set of cliques into several nonoverlapping and nonempty blocks, and then adjust clique marginals in each block locally. Thus, high-order matrix operations can be avoided and the IPS procedure is accelerated. We modify the Swendsen–Wang Algorithm and apply the simulated annealing algorithm to find an approximation to the optimal partition which leads to the least complexity. This strategy of partitioning cliques can also speed up the existing IIPS and IHT procedures. Numerical experiments are presented to demonstrate the competitive performance of our new implementations and strategies.

Learning the Structure of Mixed Graphical Models

P. 230-253

Jason D. Lee - Trevor J. Hastie

Abstract

We consider the problem of learning the structure of a pairwise graphical model over continuous and discrete variables. We present a new pairwise model for graphical models with both continuous and discrete variables that is amenable to structure learning. In previous work, authors have considered structure learning of Gaussian graphical models and structure learning of discrete models. Our approach is a natural generalization of these two lines of work to the mixed case. The penalization scheme involves a novel symmetric use of the group-lasso norm and follows naturally from a particular parameterization of the model. Supplementary materials for this article are available online.

Image Denoising by a Local Clustering Framework

P. 254-273

Partha Sarathi Mukherjee - Peihua Qiu

Abstract

Images often contain noise due to imperfections in various image acquisition techniques. Noise should be removed from images so that the details of image objects (e.g., blood vessels, inner foldings, or tumors in the human brain) can be clearly seen, and the subsequent image analyses are reliable. With broad usage of images in many disciplines—for example, medical science—image denoising has become an important research area. In the literature, there are many different types of image denoising techniques, most of which aim to preserve image features, such as edges and edge structures, by estimating them explicitly or implicitly. Techniques based on explicit edge detection usually require certain assumptions on the smoothness of the image intensity surface and the edge curves which are often invalid especially when the image resolution is low. Methods that are based on implicit edge detection often use multiresolution

smoothing, weighted local smoothing, and so forth. For such methods, the task of determining the correct image resolution or choosing a reasonable weight function is challenging. If the edge structure of an image is complicated or the image has many details, then these methods would blur such details. This article presents a novel image denoising framework based on local clustering of image intensities and adaptive smoothing. The new denoising method can preserve complicated edge structures well even if the image resolution is low. Theoretical properties and numerical studies show that it works well in various applications.

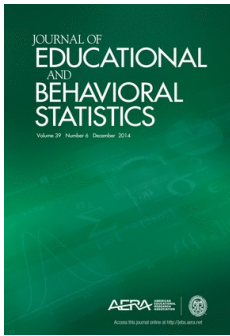
Spatially Weighted Principal Component Analysis for Imaging Classification

P. 274-296

Ruixin Guo - Mihye Ahn - Hongtu Zhu Hongtu Zhu

Abstract

The aim of this article is to develop a supervised dimension-reduction framework, called spatially weighted principal component analysis (SWPCA), for high-dimensional imaging classification. Two main challenges in imaging classification are the high dimensionality of the feature space and the complex spatial structure of imaging data. In SWPCA, we introduce two sets of novel weights, including global and local spatial weights, which enable a selective treatment of individual features and incorporation of the spatial structure of imaging data and class label information. We develop an efficient two-stage iterative SWPCA algorithm and its penalized version along with the associated weight determination. We use both simulation studies and real data analysis to evaluate the finite-sample performance of our SWPCA. The results show that SWPCA outperforms several competing principal component analysis (PCA) methods, such as supervised PCA (SPCA), and other competing methods, such as sparse discriminant analysis (SDA).



Journal of educational and behavioral statistics, ISSN 1076-9986
Volume 40, number 2 (april 2015)

Detection of Differential Item Functioning Using the Lasso Approach

P. 111-135

David Magis - Francis Tuerlinckx - Paul De Boeck

Abstract

This article proposes a novel approach to detect differential item functioning (DIF) among dichotomously scored items. Unlike standard DIF methods that perform an item-by-item analysis, we propose the “LR lasso DIF method”: logistic regression (LR) model is formulated for all item responses. The model contains item-specific intercepts, an effect of the sum score, and item-group interaction (i.e., DIF) effects, with a lasso penalty on all DIF parameters. Optimal penalty parameter selection is investigated through several known information criteria (Akaike information criterion, Bayesian information criterion, and cross validation) as well as through a newly developed alternative. A simulation study was conducted to compare the global performance of the suggested LR lasso DIF method to the LR and Mantel–Haenszel methods (in terms of false alarm and hit rates). It is concluded that for small samples, the LR lasso DIF approach globally outperforms the LR method, and also the Mantel–Haenszel method, especially in the presence of item impact, while it yields similar results with larger samples.

Weakly Informative Prior for Point Estimation of Covariance Matrices in Hierarchical Models

P. 136-157

Yeojin Chung - Andrew Gelman - Sophia Rabe-Hesketh - Jingchen Liu - Vincent Dorie

Abstract

When fitting hierarchical regression models, maximum likelihood (ML) estimation has computational (and, for some users, philosophical) advantages compared to full Bayesian inference, but when the number of groups is small, estimates of the covariance matrix (Σ) of group-level varying coefficients are often degenerate. One can do better, even from a purely point estimation perspective, by using a prior distribution or penalty function. In this article, we use Bayes modal estimation to obtain positive definite covariance matrix estimates. We recommend a class of Wishart (not inverse-Wishart) priors for Σ with a default choice of hyperparameters, that is, the degrees of freedom are set equal to the number of varying coefficients plus 2, and the scale matrix is the identity matrix multiplied by a value that is large relative to the scale of the problem. This prior is equivalent to independent gamma priors for the eigenvalues of Σ with shape parameter 1.5 and rate parameter close to 0. It is also equivalent to independent gamma priors for the variances with the same hyperparameters multiplied by a function of the correlation coefficients. With this default prior, the posterior mode for Σ is always strictly positive definite. Furthermore, the resulting uncertainty for the fixed coefficients is less underestimated than under classical ML or restricted maximum likelihood estimation. We also suggest an extension of our method that can be used when stronger prior information is available for some of the variances or correlations.

Practical Issues in Estimating Achievement Gaps From Coarsened Data

P. 158-189

Sean F. Reardon - Andrew D. Ho

Abstract

In an earlier paper, we presented methods for estimating achievement gaps when test scores are coarsened into a small

number of ordered categories, preventing fine-grained distinctions between individual scores. We demonstrated that gaps can nonetheless be estimated with minimal bias across a broad range of simulated and real coarsened data scenarios. In this article, we extend this previous work to obtain practical estimates of the imprecision imparted by the coarsening process and of the bias imparted by measurement error. In the first part of this article, we derive standard error estimates and demonstrate that coarsening leads to only very modest increases in standard errors under a wide range of conditions. In the second part of this article, we describe and evaluate a practical method for disattenuating gap estimates to account for bias due to measurement error.

An Evaluation of Empirical Bayes's Estimation of Value-Added Teacher Performance Measures

P. 190-222

Cassandra M. Guarino - Michelle Maxfield - Mark D. Reckase - Paul N. Thompson - Jeffrey M. Wooldridge

Abstract

Empirical Bayes's (EB) estimation has become a popular procedure used to calculate teacher value added, often as a way to make imprecise estimates more reliable. In this article, we review the theory of EB estimation and use simulated and real student achievement data to study the ability of EB estimators to properly rank teachers. We compare the performance of EB estimators with that of other widely used value-added estimators under different teacher assignment scenarios. We find that, although EB estimators generally perform well under random assignment (RA) of teachers to classrooms, their performance suffers under nonrandom teacher assignment. Under non-RA, estimators that explicitly (if imperfectly) control for the teacher assignment mechanism perform the best out of all the estimators we examine. We also find that shrinking the estimates, as in EB estimation, does not itself substantially boost performance.



Selection, selection, selection: the impact of return migration

P. 535-563

Jackline Wahba

Abstract

The evidence on the impact of return migration on the sending country is rather sparse, though growing. The contribution of this paper is in addressing various selectivity problems while quantifying the impact of return migration on wages of returnees using non-experimental data. Using Egyptian household-level survey data, I estimate the wages of return migrants controlling for several selectivity biases arising from emigration choice, return migration choice, labor force participation choice, and occupational choice following return. The findings provide strong evidence that overseas temporary migration results in a wage premium upon return, even after controlling for the various potential selection biases. However, the estimates underscore the significance of controlling for both emigration and return migration selections. Ignoring the double selectivity in migration would overestimate the impact of return migration on the wage premium of returnees, as migrants are positively selected relative to non-migrants, but returnees are negatively selected among migrants.

Gender discrimination in the allocation of migrant household resources

P. 565-592

Francisca M. Antman

Abstract

This paper considers the relationship between international migration and gender discrimination through the lens of decision-making power over intrahousehold resource allocation. The endogeneity of migration is addressed with a difference-in-differences style identification strategy and a model with household fixed effects. The results suggest that while a migrant household head is away, a greater share of resources is spent on girls relative to boys and his spouse commands greater decision-making power. Once the head returns home, however, a greater share of resources goes to boys, and there is suggestive evidence of greater authority for the head of household.

Siblings' interaction in migration decisions: who provides for the elderly left behind?

P. 593-629

Tobias Stöhr

Abstract

In most poor countries, with high emigration rates, elderly people are dependent on their children for the provision of care and income. This paper is the first to explicitly model and estimate social interaction between siblings' migration decisions in such settings. The interaction consists of two effects with opposite signs; a chain migration effect that can cause traditional caregiving structures to break down and an opposing specialization effect that increases family members' incentives to remain at home and provide care when their siblings migrate. The estimates for Moldova, one of the countries with the highest emigration rates in the world, indicate that siblings' interaction strongly decreases their equilibrium emigration rates. Siblings' interaction is found to increase in line with the incentives that are assumed in the model. Hence, the paper provides evidence of the robustness of families' informal security arrangements to large-scale emigration and has important implications for policies that aim at the population left behind.

Migration and young child nutrition: evidence from rural China

P. 631-657

Ren Mu - Alan de Brauw

Abstract

The unprecedented large-scale rural-to-urban migration in China has left many rural children living apart from their parents. In this study, we examine the impact of parental migration on the nutritional status of young children in rural areas. We use the interaction terms between wage growth, by gender, in provincial capital cities and initial village migrant networks as instrumental variables to account for migration selection. Our results show that parental migration has no significant effect on the height of children, but it improves their weight. We provide suggestive evidence that the improvement in weight may be achieved through increased access to tap water in migrant households.

How do the foreign-born perform in inventive activity? Evidence from Sweden

P. 659-695

Yannu Zheng - Olof Ejermo

Abstract

Using a new database that matches patent applications by Swedish residents with demographic information from 1985 to 2007, we examine differences in inventive performance by individuals of foreign and domestic origins, in terms of quantity (probability of patenting, total number of patents per inventor) and quality (forward citations, probability of grant) of patents. We further compare adult and child immigrants with their Swedish-born counterparts. Holding other variables constant, we find that the immigrants are generally less likely to patent than the Swedish-born. Nonetheless, the general group of immigrant inventors, including those who migrated as adults, performs as well as the native inventors and therefore seems more positively selected. Compared with the Swedish-born, the immigrants who migrated as children are disadvantaged in both quantity and quality of patents, which may be linked to a lack of Sweden-specific human capital. Whether education was received in Sweden does not seem to make a difference for the immigrants who migrated as adults. In summary, this study provides an initial impression of the inventive performance, contribution and challenges of distinct groups of immigrants who have differing characteristics and backgrounds.

Immigration and crime: evidence from victimization data

P. 697-736

Luca Nunziata

Abstract

We exploit the increase in immigration flows into western European countries that took place in the 2000s to assess whether immigration affects crime victimization and the perception of criminality among European natives. Using data from the European Social Survey, the Labour Force Survey and other sources, we provide a set of fixed effects and instrumental variable estimations that deal with the endogenous sorting of immigration by region and with the sampling error in survey-based measures of regional immigration shares, whose implications in terms of attenuation bias are investigated by means of Monte Carlo simulations. Our empirical findings show that an increase in immigration does not affect crime victimization, but it is associated with an increase in the fear of crime, the latter being consistently and positively correlated with the natives' unfavourable attitude toward immigrants. Our results reveal a misconception of the link between immigration and crime among European natives.

School and neighborhood: residential location choice of immigrant parents in the Los Angeles Metropolitan area

P. 737-783

Crystal Zhan

Abstract

This paper studies how immigrant parents value education for their children in the United States when making residential decisions. Parent valuation of education is examined through the differential effects of school quality on the residential location choices of households with and without children. The analysis relies on data from the 2000 Census and focuses on the Los Angeles Metropolitan Area. The results suggest that immigrant parents place a positive weight on school quality when choosing residences. The weight assigned to school is positively associated with household

income and householder's education. The paper further explores variation across immigrants to get at the potential economic mechanisms for differential valuation of school quality. Number of school-aged children in the household, selective migration, and potential returns to education may explain variation in the emphasis immigrant parents place on school quality in residential location choices.

Assimilation in multilingual cities

P. 785-815

Javier Ortega - Gregory Verdugo

Abstract

We characterise how the assimilation patterns of minorities into the strong and the weak language differ in a situation of asymmetric bilingualism. Using large variations in language composition in Canadian cities from the 2001 and 2006 Censuses, we show that the differences in the knowledge of English by immigrant allophones (i.e., the immigrants with a mother tongue other than English and French) in English-majority cities are mainly due to sorting across cities. Instead, in French-majority cities, learning plays an important role in explaining differences in knowledge of French. In addition, the presence of large anglophone minorities deters much more the assimilation into French than the presence of francophone minorities deters the assimilation into English. Finally, we find that language distance plays a much more important role in explaining assimilation into French, and that assimilation into French is much more sensitive to individual characteristics than assimilation into English. Some of these asymmetric assimilation patterns extend to anglophone and francophone immigrants, but no evidence of learning is found in this case.

Life satisfaction of immigrants: does cultural assimilation matter?

P. 817-844

Viola Angelini - Laura Casi - Luca Corazzini

Abstract

To investigate empirically the association between a direct measure of assimilation with a host culture and immigrants' subjective well-being, this study uses data from the German Socio-Economic Panel. A positive, significant association arises between cultural assimilation and immigrants' life satisfaction, even after controlling for several potential confounding factors, such as immigrants' individual (demographic and socio-economic) characteristics and regional controls that capture their external social conditions. Finally, the strength of the association varies with time since migration; it is significant for "established" and second-generation immigrants but vanishes for "recent" immigrants.

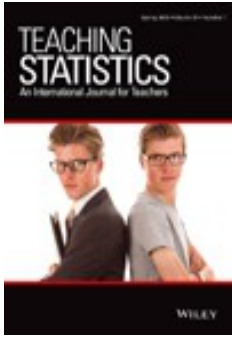
Genetic distance, immigrants' identity, and labor market outcomes

P. 845-868

Asadul Islam - Paul A. Raschky

Abstract

Consistent estimates of the effect of immigrants' identity on labor market outcomes is complicated by the endogenous relationship between performance on the labor market and attitudes towards ethnic identity. This paper uses measures of genetic distance between immigrants' home and host countries as instruments for immigrants' identity. We find some evidence for adverse effects of home country identity on male immigrants' unemployment likelihood. Our results also suggest that a stronger host country identity only has a systematic effect on employment and job satisfaction. Overall, immigrants' identity appears to play only a negligible role in immigrants' labor market performance. Our analysis also shows the importance of accounting for endogeneity in the relationship between immigrants' ethnic identity and labor market outcomes.



Teaching statistics, ISSN 0141-982X
Volume 37, number 2 (summer 2015)

Did the Gamemakers Fix the Lottery in the Hunger Games?

P. 37-40

Kyle Caudle - Erica Daniels

Abstract

The Hunger Games is an annual event in the fictional country of Panem. Each year, 24 children (tributes) are chosen by lottery from 12 districts to fight to the death in the arena for the entertainment of the Capitol citizens. Using statistical analysis and computer simulations, we will explore the possibility that the Gamemakers, those in charge of planning the Hunger Games, fixed the lottery. Using the fictitious data from Suzanne Collins' book *the Hunger Games*, we show how students can learn how to perform a permutation goodness of fit test.

Motivating Inquiry in Statistics and Probability in the Primary Classroom

P. 41-47

Aisling Leavy - Mairéad Hourigan

Abstract

We describe how the use of a games environment combined with technology supports upper primary children in engaging with a concept traditionally considered too advanced for the primary classes: *The Law of Large Numbers*.

Active Learning and Threshold Concepts in Multiple Testing that can Further Develop Student Critical Statistical Thinking

P. 48-53

Desley White

Abstract

Two practical activities are described, which aim to support critical thinking about statistics as they concern multiple outcomes testing. Formulae are presented in Microsoft Excel spreadsheets, which are used to calculate the inflation of error associated with the quantity of tests performed. This is followed by a decision-making exercise, where an Excel calculator is used to adjust for multiple outcomes prior to deciding whether or not an intervention has been successful.

Numbers defy the law of large numbers

P. 54-60

Ruma Falk - Avital Lavie Lann

Abstract

As the number of independent tosses of a fair coin grows, the rates of heads and tails tend to equality. This is misinterpreted by many students as being true also for the absolute numbers of the two outcomes, which, conversely, depart unboundedly from each other in the process. Eradicating that misconception, as by coin-tossing experiments, should be incorporated early on into learning the law of large numbers.

Are we able to pass the mission of statistics to students?

Richard Hindls - Stanislava Hronová

P. 61-65

Abstract

The article illustrates our long term experience in teaching statistics for non-statisticians, especially for students of economics and humanities. The article is focused on some problems of the basic course that can weaken the interest in statistics or lead to false use of statistic methods.
