



Biblioteca del Instituto de Estadística y Cartografía de Andalucía

Resúmenes de revistas
Junio 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 **junio 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.



Biometrika, ISSN 0006-3444
Volume 102, number 2 (june 2015)

Testing differential networks with applications to the detection of gene-gene interactions

P. 247-266

Yin Xia - Tianxi Cai - T. Tony Cai

Abstract

Model organisms and human studies have yielded increasing empirical evidence that interactions among genes contribute broadly to genetic variation of complex traits. In the presence of gene-gene interactions, the dimensionality of the feature space becomes extremely high relative to the sample size. This poses a significant methodological challenge in the identification of gene-gene interactions. In this paper, by using a Gaussian graphical model framework, we translate the problem of identifying gene-gene interactions associated with a binary trait D into an inference problem on the difference of two high-dimensional precision matrices that summarize the conditional dependence network structures of the genes. We propose a procedure for testing the differential network globally, which is particularly powerful against sparse alternatives. In addition, a multiple testing procedure with false discovery rate control is developed to infer the specific structure of the differential network. Theoretical justification is provided to ensure the validity of the proposed tests, and optimality results are derived under sparsity assumptions. Through a simulation study we demonstrate that the proposed tests maintain the desired error rates under the null hypothesis and have good power under the alternative hypothesis. The methods are applied to a breast cancer gene expression study.

Hierarchical recognition of sparse patterns in large-scale simultaneous inference

P. 267 -280

Wenguang Sun -Zhi Wei

Abstract

We study how to separate signals from noisy data accurately and determine the patterns of the selected signals. Controlling the inflation of false positive errors is important in large-scale simultaneous inference but has not been addressed in the pattern recognition literature. We develop a decision-theoretic framework and formulate the sparse pattern recognition problem as a simultaneous inference problem with multiple decision trees. Oracle and adaptive classifiers are proposed for maximizing the expected number of true positives subject to a constraint on the overall false positive rate. Existing results on multiple testing are extended by allowing more than two states of nature, hierarchical decision-making and new error rate concepts.

On random-effects meta-analysis

P. 281-294

D. Zeng -D. Y. Lin

Abstract

Meta-analysis is widely used to compare and combine the results of multiple independent studies. To account for between-study heterogeneity, investigators often employ random-effects models, under which the effect sizes of interest are assumed to follow a normal distribution. It is common to estimate the mean effect size by a weighted linear combination of study-specific estimators, with the weight for each study being inversely proportional to the sum of the variance of the effect-size estimator and the estimated variance component of the random-effects distribution. Because the estimator of the variance

component involved in the weights is random and correlated with study-specific effect-size estimators, the commonly adopted asymptotic normal approximation to the meta-analysis estimator is grossly inaccurate unless the number of studies is large. When individual participant data are available, one can also estimate the mean effect size by maximizing the joint likelihood. We establish the asymptotic properties of the meta-analysis estimator and the joint maximum likelihood estimator when the number of studies is either fixed or increases at a slower rate than the study sizes and we discover a surprising result: the former estimator is always at least as efficient as the latter. We also develop a novel resampling technique that improves the accuracy of statistical inference. We demonstrate the benefits of the proposed inference procedures using simulated and empirical data.

Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator

P. 295-313

A. Doucet - M. K. Pitt - G. Deligiannidis - R. Kohn

Abstract

When an unbiased estimator of the likelihood is used within a Metropolis–Hastings chain, it is necessary to trade off the number of Monte Carlo samples used to construct this estimator against the asymptotic variances of the averages computed under this chain. Using many Monte Carlo samples will typically result in Metropolis–Hastings averages with lower asymptotic variances than the corresponding averages that use fewer samples; however, the computing time required to construct the likelihood estimator increases with the number of samples. Under the assumption that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with variance inversely proportional to the number of samples and independent of the parameter value at which it is evaluated, we provide guidelines on the number of samples to select. We illustrate our results by considering a stochastic volatility model applied to stock index returns.

A useful variant of the Davis–Kahan theorem for statisticians

P. 315-323

Y. Yu - T. Wang - R. J. Samworth

Abstract

The Davis–Kahan theorem is used in the analysis of many statistical procedures to bound the distance between subspaces spanned by population eigenvectors and their sample versions. It relies on an eigenvalue separation condition between certain population and sample eigenvalues. We present a variant of this result that depends only on a population eigenvalue separation condition, making it more natural and convenient for direct application in statistical contexts, and provide an improvement in many cases to the usual bound in the statistical literature. We also give an extension to situations where the matrices under study may be asymmetric or even non-square, and where interest is in the distance between subspaces spanned by corresponding singular vectors.

Information-theoretic optimality of observation-driven time series models for continuous responses

P. 325-343

F. Blasques - S. J. Koopman - A. Lucas

Abstract

We investigate information-theoretic optimality properties of the score function of the predictive likelihood as a device for updating a real-valued time-varying parameter in a univariate observation-driven model with continuous responses. We restrict our attention to models with updates of one lag order. The results provide theoretical justification for a class of score-driven models which includes the generalized autoregressive conditional heteroskedasticity model as a special case. Our main contribution is to show that only parameter updates based on the score will always reduce the local Kullback–Leibler divergence between the true conditional density and the model-implied conditional density. This result holds irrespective of the severity of model misspecification. We also show that use of the score leads to a considerably smaller global Kullback–Leibler divergence in empirically relevant settings. We illustrate the theory with an application to time-varying volatility models. We show that the reduction in Kullback–Leibler divergence across a range of different settings can be substantial

compared to updates based on, for example, squared lagged observations.

On the dependence structure of bivariate recurrent event processes: inference

P. 345-358

Jing Ning - Yong Chen - Chunyan Cai - Xuelin Huang - Mei-Cheng Wang

Abstract

Bivariate or multivariate recurrent event processes are often encountered in longitudinal studies in which more than one type of event is of interest. There has been much research on regression analysis for such data, but little has been done to measure the dependence between recurrent event processes. We propose a time-dependent measure, termed the rate ratio, to assess the local dependence between two types of recurrent event processes. We model the rate ratio as a parametric function of time, and leave unspecified all other aspects of the distribution. We develop a composite likelihood procedure for model fitting and parameter estimation. We show that the proposed estimator is consistent and asymptotically normal. Its finite sample performance is evaluated by simulation and illustrated by an application to a soft tissue sarcoma study.

A Möbius transformation-induced distribution on the torus

P. 359-370

Shogo Kato - Arthur Pewsey

Abstract

We propose a five-parameter bivariate wrapped Cauchy distribution as a unimodal model for toroidal data. It is highly tractable, displays numerous desirable properties, including marginal and conditional distributions that are all wrapped Cauchy, and arises as an appealing submodel of a six-parameter distribution obtained by applying Möbius transformation to a pre-existing bivariate circular model. Method of moments and maximum likelihood estimation of its parameters are fast, and tests for independence and goodness-of-fit are available. An analysis involving dihedral angles of the proteinogenic amino acid Tyrosine illustrates the distribution's application. A Markov process for circular data is also explored.

Maximum projection designs for computer experiments

P. 371-380

V. Roshan Joseph - Evren Gul - Shan Ba

Abstract

Space-filling properties are important in designing computer experiments. The traditional maximin and minimax distance designs consider only space-filling in the full-dimensional space; this can result in poor projections onto lower-dimensional spaces, which is undesirable when only a few factors are active. Restricting maximin distance design to the class of Latin hypercubes can improve one-dimensional projections but cannot guarantee good space-filling properties in larger subspaces. We propose designs that maximize space-filling properties on projections to all subsets of factors. We call our designs maximum projection designs. Our design criterion can be computed at no more cost than a design criterion that ignores projection properties.

Automatic structure recovery for additive models

P. 381-395

Yichao Wu - Leonard A. Stefanski

Abstract

We propose an automatic structure recovery method for additive models, based on a backfitting algorithm coupled with local polynomial smoothing, in conjunction with a new kernel-based variable selection strategy. Our method produces estimates of the set of noise predictors, the sets of predictors that contribute polynomially at different degrees up to a specified degree M , and the set of predictors that contribute beyond polynomially of degree M . We prove consistency of the proposed method, and describe an extension to partially linear models. Finite-sample performance of the method is illustrated via Monte Carlo studies and a real-data example.

Jump information criterion for statistical inference in estimating discontinuous curves

P. 397-408

Zhiming Xia - Peihua Qiu

Abstract

Nonparametric regression analysis when the regression function is discontinuous has many applications. Existing methods for estimating a discontinuous regression curve usually assume that the number of jumps in the regression curve is known beforehand, which is unrealistic in some situations. Although there has been research on estimation of a discontinuous regression curve when the number of jumps is unknown, the problem remains mostly open because such research often requires assumptions on other related quantities, such as a known minimum jump size. In this paper we propose a jump information criterion which consists of a term measuring the fidelity of the estimated regression curve to the observed data and a penalty related to the number of jumps and the jump sizes. The number of jumps can then be determined by minimizing our criterion. Theoretical and numerical studies show that our method works well.

A validated information criterion to determine the structural dimension in dimension reduction models

P. 409-420

Yanyuan Ma - Xinyu Zhang

Abstract

A crucial component of performing sufficient dimension reduction is to determine the structural dimension of the reduction model. We propose a novel information criterion-based method for this purpose, a special feature of which is that when examining the goodness-of-fit of the current model, one needs to perform model evaluation by using an enlarged candidate model. Although the procedure does not require estimation under the enlarged model of dimension $k+1$, the decision as to how well the current model of dimension k fits relies on the validation provided by the enlarged model; thus we call this procedure the validated information criterion, $vic(k)$. Our method is different from existing information criterion-based model selection methods; it breaks free from dependence on the connection between dimension reduction models and their corresponding matrix eigenstructures, which relies heavily on a linearity condition that we no longer assume. We prove consistency of the proposed method, and its finite-sample performance is demonstrated numerically.

Effective dimension reduction for sparse functional data

P. 421-437

F. Yao - E. Lei - Y. Wu

Abstract

We propose a method of effective dimension reduction for functional data, emphasizing the sparse design where one observes only a few noisy and irregular measurements for some or all of the subjects. The proposed method borrows strength across the entire sample and provides a way to characterize the effective dimension reduction space, via functional cumulative slicing. Our theoretical study reveals a bias-variance trade-off associated with the regularizing truncation and decaying structures of the predictor process and the effective dimension reduction space. A simulation study and an application illustrate the superior finite-sample performance of the method.

Envelopes and reduced-rank regression

P. 439-456

R. Dennis Cook - Liliana Forzani - Xin Zhang

Abstract

We incorporate the nascent idea of envelopes (Cook et al., *Statist. Sinica* 20, 927–1010) into reduced-rank regression by proposing a reduced-rank envelope model, which is a hybrid of reduced-rank and envelope regressions. The proposed model has total number of parameters no more than either of reduced-rank regression or envelope regression. The resulting estimator is at least as efficient as both existing estimators. The methodology of this paper can be adapted to other envelope models, such as partial envelopes (Su & Cook, *Biometrika* 98, 133–46) and

envelopes in predictor space (Cook et al., *J. R. Statist. Soc. B* 75, 851–77).

On the degrees of freedom of reduced-rank estimators in multivariate regression

P. 457-477

A. Mukherjee - K. Chen - N. Wang - J. Zhu

Abstract

We study the effective degrees of freedom of a general class of reduced-rank estimators for multivariate regression in the framework of Stein's unbiased risk estimation. A finite-sample exact unbiased estimator is derived that admits a closed-form expression in terms of the thresholded singular values of the least-squares solution and hence is readily computable. The results continue to hold in the high-dimensional setting where both the predictor and the response dimensions may be larger than the sample size. The derived analytical form facilitates the investigation of theoretical properties and provides new insights into the empirical behaviour of the degrees of freedom. In particular, we examine the differences and connections between the proposed estimator and a commonly-used naive estimator. The use of the proposed estimator leads to efficient and accurate prediction risk estimation and model selection, as demonstrated by simulation studies and a data example.

Effective degrees of freedom: a flawed metaphor

P. 479-485

Lucas Janson - William Fithian - Trevor J. Hastie

Abstract

To most applied statisticians, a fitting procedure's degrees of freedom is synonymous with its model complexity, or its capacity for overfitting to data. In particular, the degrees of freedom is often used to parameterize the bias-variance trade-off in model selection. We argue that, on the contrary, model complexity and degrees of freedom may correspond very poorly. We exhibit and theoretically explore various fitting procedures for which the degrees of freedom is not monotonic in the model complexity parameter and can exceed the total dimension of the ambient space even in very simple settings. We show that the degrees of freedom for any nonconvex projection method can be unbounded.

Semiparametric exponential families for heavy-tailed data

P. 486-493

William Fithian - Stefan Wager

Abstract

We propose a semiparametric method for fitting the tail of a heavy-tailed population given a relatively small sample from that population and a larger sample from a related background population. We model the tail of the small sample as an exponential tilt of the better-observed large-sample tail, using a robust sufficient statistics motivated by extreme value theory. In particular, our method induces an estimator of the small-population mean, and we give theoretical and empirical evidence that this estimator outperforms methods that do not use the background sample. We demonstrate substantial efficiency gains over competing methods in simulation and on data from a large controlled experiment conducted by Facebook.

Optimum designs for two treatments with unequal variances in the presence of covariates

P. 494-499

A. C. Atkinson

Abstract

Optimum designs are described for two treatments with different variances when covariates are included in the model. The design, a generalization of Neyman allocation, are required in personalized medicine to model the effect of covariates on the choice of treatment. The use of the designs in clinical trials is indicated. D-optimality of the designs is established using results from Kiefer's general equivalence theorem. The results are obtained with the use of surprisingly elementary algebra.



Journal of agricultural, biological, and environmental statistics,
ISSN 1085-7117
Volume 20, number 1 (march 2015)

Estimating Abundance from Counts in Large Data Sets of Irregularly Spaced Plots using Spatial Basis Functions

P. 1-27

Jay M. Ver Hoef - John K. Jansen

Abstract

Monitoring plant and animal populations is an important goal for both academic research and management of natural resources. Successful management of populations often depends on obtaining estimates of their mean or total over a region. The basic problem considered in this paper is the estimation of a total from a sample of plots containing count data, but the plot placements are spatially irregular and non-randomized. Our application had counts from thousands of irregularly spaced aerial photo images. We used change-of-support methods to model counts in images as a realization of an inhomogeneous Poisson process that used spatial basis functions to model the spatial intensity surface. The method was very fast and took only a few seconds for thousands of images. The fitted intensity surface was integrated to provide an estimate from all unsampled areas, which is added to the observed counts. The proposed method also provides a finite area correction factor to variance estimation. The intensity surface from an inhomogeneous Poisson process tends to be too smooth for locally clustered points, typical of animal distributions, so we introduce several new overdispersion estimators due to poor performance of the classic one. We used simulated data to examine estimation bias and to investigate several variance estimators with overdispersion. A real example is given of harbor seal counts from aerial surveys in an Alaskan glacial fjord.

Analysing Mark–Recapture–Recovery Data in the Presence of Missing Covariate Data Via Multiple Imputation

P. 28-46

Hannah Worthington - Ruth King

Abstract

We consider mark–recapture–recovery data with additional individual time-varying continuous covariate data. For such data it is common to specify the model parameters, and in particular the survival probabilities, as a function of these covariates to incorporate individual heterogeneity. However, an issue arises in relation to missing covariate values, for (at least) the times when an individual is not observed, leading to an analytically intractable likelihood. We propose a two-step multiple imputation approach to obtain estimates of the demographic parameters. Firstly, a model is fitted to only the observed covariate values. Conditional on the fitted covariate model, multiple “complete” datasets are generated (i.e. all missing covariate values are imputed). Secondly, for each complete dataset, a closed form complete data likelihood can be maximised to obtain estimates of the model parameters which are subsequently combined to obtain an overall estimate of the parameters. Associated standard errors and 95% confidence intervals are obtained using a non-

parametric bootstrap. A simulation study is undertaken to assess the performance of the proposed two-step approach. We apply the method to data collected on a well-studied population of Soay sheep and compare the results with a Bayesian data augmentation approach. Supplementary materials accompanying this paper appear on-line.

Bayesian Nonparametric Models of Circular Variables Based on Dirichlet Process Mixtures of Normal Distributions

P. 47-64

Gabriel Nuñez-Antonio

Abstract

This article introduces two new Bayesian nonparametric models for circular data based on Dirichlet process (DP) mixtures of normal distributions. The first model is a projected DP mixture of bivariate normals and the second approach is based on a wrapped DP mixture of normal distributions. We show how to carry out inference for these models based on a slice sampling scheme and introduce an approach to estimating a variant of the deviance information criterion which is appropriate in the context of latent variable models. Our models are then compared with both simulated and real data examples.

Assessing Assay Variability of Pesticide Metabolites in the Presence of Heavy Left-Censoring

P. 65-82

Haiying Chen - Sara A. Quandt - Dana Boyd Barr

Abstract

Assessing assay variability for field samples in environmental research is challenging, since a quantitative assay is typically constrained by a lower limit of detection. The purpose of this paper is to compare three parametric models for assessing assay variability using duplicate data subject to heavy left-censoring. Efron information criterion (EIC) and Bayesian information criterion (BIC) are used to aid in model selections. Distributional parameter estimates are obtained using maximum likelihood estimation for bivariate lognormal, bivariate zero-inflated lognormal, and bivariate 3-component mixture models. We illustrate a practical application using duplicate pesticide data from the Community Participatory Approach to Measuring Farmworker Pesticide Exposure (PACE3) study. Furthermore, a simulation study is conducted to empirically evaluate the performance of the three models. The results from PACE3 indicate that the bivariate zero-inflated lognormal model is fairly competitive based on EIC or BIC. Further, total variability for the lognormal component can be decomposed into between-subject and within-subject variance based on this model. Assay variability estimates such as within-subject coefficient variation, minimum detectable change, and probability of k -fold difference can be easily derived under the bivariate zero-inflated lognormal model. Additionally, the assay variability is rather large for the PACE3 data. Therefore, apparent longitudinal change in pesticide exposure should be examined cautiously in the context of substantial assay variability. Supplementary materials accompanying this paper appear online.

Maximum Pairwise Pseudo-likelihood Estimation of the Covariance Matrix from Left-Censored Data

P. 83-99

Michael P. Jones - Sarah S. Perry

Abstract

Toxicological studies often depend on laboratory assays that have thresholds below which environmental pollutants cannot be measured with accuracy. Exposure levels below this limit of detection may well be toxic and hence it is vital to use data analytic methods that handle such left-censored data with as little estimation bias as possible. In an

on-going study for which our methodology is developed, levels of residential exposure to polychlorinated biphenyls (PCBs) and the interrelationships of their subtypes (congeners) are characterized. In any given sample many of the congeners may fall below the detection limit. The main problem tackled in this paper is estimation of mean exposure levels and corresponding covariance and correlation matrices for a large number of potentially left-censored measures that have very low bias and are computationally feasible. The proposed methods are likelihood based, using marginal likelihoods for means and variances and pairwise pseudo-likelihoods for correlations and covariances. In the simple bivariate case, head-to-head comparisons show the proposed methods to be computationally more stable than ordinary maximum likelihood estimates (MLEs) and still maintain comparable bias. When the number of variables is much larger than 2, the proposed methods are far more computationally feasible than MLE. Furthermore, they exhibit much less bias when compared to popular imputation procedures. Analysis of the PCB data uncovered interesting correlational structures.

Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting

P. 100-120

Caroline Carrico - Chris Gennings

Abstract

In risk evaluation, the effect of mixtures of environmental chemicals on a common adverse outcome is of interest. However, due to the high dimensionality and inherent correlations among chemicals that occur together, the traditional methods (e.g. ordinary or logistic regression) suffer from collinearity and variance inflation, and shrinkage methods have limitations in selecting among correlated components. We propose a weighted quantile sum (WQS) approach to estimating a body burden index, which identifies “bad actors” in a set of highly correlated environmental chemicals. We evaluate and characterize the accuracy of WQS regression in variable selection through extensive simulation studies through sensitivity and specificity (i.e., ability of the WQS method to select the bad actors correctly and not incorrect ones). We demonstrate the improvement in accuracy this method provides over traditional ordinary regression and shrinkage methods (lasso, adaptive lasso, and elastic net). Results from simulations demonstrate that WQS regression is accurate under some environmentally relevant conditions, but its accuracy decreases for a fixed correlation pattern as the association with a response variable diminishes. Nonzero weights (i.e., weights exceeding a selection threshold parameter) may be used to identify bad actors; however, components within a cluster of highly correlated active components tend to have lower weights, with the sum of their weights representative of the set.

Robust Joint Non-linear Mixed-Effects Models and Diagnostics for Censored HIV Viral Loads with CD4 Measurement Error

P. 121-139

Dipankar Bandyopadhyay - Luis M. Castro

Abstract

Despite technological advances in efficiency enhancement of quantification assays, biomedical studies on HIV RNA collect viral load responses that are often subject to detection limits. Moreover, some related covariates such as CD4 cell count may be often measured with errors. Censored non-linear mixed-effects models are routinely used to analyze this type of data and are based on normality assumptions for the between-subject and within-subject random terms. However, derived inference may not be robust when the underlying normality assumptions are questionable, especially in presence of skewness and heavy tails. In this article, we address these issues simultaneously under a Bayesian paradigm through joint modeling of the response and covariate processes using an attractive class of skew-normal independent densities. The methodology is illustrated using a case study on longitudinal HIV viral loads. Diagnostics for outlier detection is immediate from the MCMC output. Both simulation and real data analysis reveal the advantage of the

proposed models in providing robust inference under non-normality situations commonly encountered in HIV/AIDS or other clinical studies.

Estimation of General Multistage Models From Cohort Data

P. 140-155

Perry de Valpine - Jonas Knape

Abstract

Many systems involve progression through a series of distinct stages, such as disease or developmental stages. In ecological studies, often individuals such as small arthropods cannot be marked, so data are collected on cohort development. Multistage models for unmarked cohort data use a distribution for each stage duration and possibly stage-specific mortality rates. We generalize previous models and present computational methods for smoothed maximum likelihood estimation. The general model allows arbitrary distribution assumptions, stage-specific mortality, unobserved stages, and correlations between stage durations using Gaussian copulas. Monte Carlo integration of the stage distributions is used to approximate the probabilities needed for the likelihood. We establish a heuristic smoothing step for the simulated probabilities that yields a smooth approximate likelihood surface. For the case of classic grasshopper cohort data, we demonstrate AIC model selection to determine which among past arbitrary constraints are actually justified by the data. Finally, we demonstrate how estimates of stage distribution parameters depend on the unknown stage correlations.

Pseudo-likelihood Estimation of Multivariate Normal Parameters in the Presence of Left-Censored Data

P. 183-204

Heather J. Hoffman - Robert E. Johnson

Abstract

Environmental data often include left-censored values reported to be less than some limit of detection (LOD). While simple imputation of a specific value such as LOD/2 is common, maximum likelihood methods accounting for censoring provide alternate ways of analyzing such data. Concentration levels of contaminants in water, for example, are typically modeled with normal or lognormal distributions. Corresponding maximum likelihood estimates (MLEs) of means and variances in univariate analyses can be obtained from standard software packages; however, multivariate analyses may be more appropriate when multiple measurements come from the same entity. For example, measures of several dissolved trace metals may be derived from freshwater stream samples. In less-polluted areas, one or more of these measures fall below the LOD. An index of overall contamination may be formed as a linear function of these measures. The desire to estimate this index led to the need to estimate the parameters in the presence of nondetects, which led to our proposed method. We propose a pseudo-likelihood method utilizing pairs of variables that provides MLEs of mean and unstructured covariance parameters corresponding to a multivariate normal or lognormal distribution in the presence of left-censored data. In conducting hypothesis tests and estimating functions of MLEs with standard errors, we apply this method to trace metal concentration data collected from freshwater streams across Virginia.



Papeles de economía española, ISSN 0210-9107
Número 144 (2015): "La industria española: un impulso necesario"

La industria española: desde la crisis hacia la fortaleza

P. 2-23

María José Moral – Consuelo Pazó

Resumen

El objetivo de este artículo es proporcionar una visión panorámica de la situación de la industria en España y su evolución reciente. La crisis en la que ha estado inmersa la economía en los últimos años hace prioritario identificar las debilidades y las fortalezas de un sector con un importante efecto arrastre sobre otras actividades económicas. Por ello, se analiza la evolución del VAB, el empleo y la productividad del trabajo, y se señalan las heterogeneidades sectoriales y territoriales. Además, se examinan las estrategias empresariales respecto a la participación en los mercados exteriores, en actividades de I+D+I y en la utilización de trabajo cualificado. Los resultados indican un incremento en estas actividades en los años finales de recesión como consecuencia, en parte, de que las empresas implicadas en estas estrategias están sobreviviendo a la crisis en mayor medida

La industria europea: retos y perspectivas

P. 24-41

José Antonio Cuenca – Esther Gordo

Resumen

Durante los años recientes, la industria europea ha tenido que hacer frente a importantes desafíos asociados a la aparición de nuevos competidores en la escena internacional y al desarrollo de nuevas formas de producción derivadas del surgimiento y extensión de las cadenas de producción global. Con el propósito de evaluar las posibilidades de que en este nuevo contexto se produzca un proceso de reindustrialización en la UEM, en este artículo se revisa la evolución reciente de las manufacturas europeas y las transformaciones estructurales que han tenido lugar a nivel sectorial. También se analizan la magnitud y principales características del proceso de inserción de las economías europeas en las cadenas de valor mundiales. A la luz de la evidencia presentada, las conclusiones de este artículo discuten las dificultades a que se enfrenta el proceso de reindustrialización europeo y la necesidad de que este se asiente en el fomento de la innovación y el capital humano.

La desindustrialización de España en el contexto europeo

P. 42-55

José Carlos Fariñas – Ana Martín Marcos – Francisco J. Velázquez

Resumen

En este artículo se examina la evolución de la industria española durante las cuatro últimas décadas en el contexto de la industria europea. La pérdida de peso relativo de la industria es común al conjunto de países de la OCDE. Desde el año 2000, España está entre el grupo de países que más reducen el peso relativo de su industria. Se examinan tres explicaciones distintas pero no excluyentes de esta pérdida de peso: el cambio estructural asociado con la evolución de la productividad y demanda relativas, el comercio y la competencia exterior, y los fenómenos de externalización y terciarización de la industria.

La manufactura española en la economía de mercado: 1980-2013. Análisis y perspectivas

P. 56-72

Vicente Salas Fumás

Resumen

Este trabajo analiza la actividad, productividad, precios, integración vertical y competitividad que explican la evolución del peso relativo de la Manufactura en el conjunto de la economía de mercado en España entre 1980 y 2013. Los resultados indican que la entrada en el euro cambia algunas relaciones estructurales de la Manufactura con el resto de sectores, y cambia también la dinámica de relación entre costes y precios de importaciones y exportaciones, con efectos negativos para el empleo manufacturero. Más empleo en la Manufactura (reindustrialización) en España y en cualquier país, pasa por que la producción nacional gane cuota en el mercado mundial de bienes, suficiente para compensar la tendencia a un empleo manufacturero menguante debido a la baja elasticidad precio y renta de la demanda de bienes, y al mayor crecimiento de la productividad en la manufactura que en el resto de sectores.

Tamaño y dinámica empresarial en la industria española: efecto de la gran recesión

P. 74-91

Xulia González

Resumen

El objetivo de este trabajo es analizar el impacto de la actual recesión sobre la dinámica empresarial en términos de entradas y salidas de empresas y creación y destrucción de empleo. Los datos empleados en el trabajo provienen fundamentalmente de dos fuentes: el Directorio Central de Empresas (DIRCE) y la Encuesta de Estrategias Empresariales (ESEE), y se centran en el sector industrial. Los resultados obtenidos muestran el fuerte proceso de destrucción neta de empresas y de empleo que ha tenido lugar desde el inicio de la gran recesión económica, concentrado de forma muy intensa en el año 2009. Por una parte, se muestra que la tasa de apertura de empresas es más cíclica que la tasa de cierres. Por otra, el fuerte ajuste de empleo que tuvo lugar en el año 2009 ha contrarrestado buena parte del crecimiento empresarial experimentado en la década precedente. Por último, encontramos que la innovación y la exportación han contribuido significativamente al crecimiento de las empresas en los años de recesión.

Las empresas manufactureras en las cadenas de valor globales: evidencia para España

P. 92-104

José Carlos Fariñas – Ana Martín Marcos – Francisco J. Velázquez

Resumen

El artículo analiza el fenómeno de la fragmentación de los procesos productivos poniendo el foco en el outsourcing y el offshoring de las manufacturas españolas. Con datos de la Encuesta sobre Estrategias Empresariales estimamos que alrededor del 40 por 100 de las empresas hacen offshoring y la intensidad es casi del 30 por 100, presentando una gran heterogeneidad sectorial. En términos relativos respecto a otros países de la OCDE, el volumen de comercio y la proporción de empresas integradas en las llamadas cadenas de valor globales son bajos. También se examina la relación entre la productividad y las estrategias de abastecimiento de inputs intermedios.

Markups, exportaciones e I + D: evidencia para el sector manufacturero en España

P. 105-122

Juan A. Mañez – María E. Rochina-Barrachina – Juan A. Sanchis-Llopis

Resumen

Este trabajo analiza conjuntamente la relación entre las actividades de exportación y de I + D por parte de las empresas y sus markups. Para ello utilizamos la metodología de estimación de markups de De Loecker y Warzynski (2012). Los datos utilizados provienen de la Encuesta sobre Estrategias Empresariales, 1993-2009. Los resultados revelan una relación positiva tanto de la actividad solo exportadora como del ejercicio conjunto de esta actividad con la I + D sobre los markups. Además, los años de crisis parecen apuntar hacia una relevancia creciente de las actividades de innovación a la hora de justificar

Capital humano y productividad: un análisis de la empresa manufacturera en España desde principios de los noventa

P. 123-138

Miguel Cardoso – Mónica Correa-López – Rafael Doménech

Resumen

Este artículo tiene como objetivos identificar algunas de las regularidades empíricas que caracterizan la productividad de las empresas en el sector manufacturero español y profundizar en el estudio de los factores candidatos a explicar su evolución de los últimos veinticinco años, con especial atención al papel del capital humano. Adoptando un enfoque microeconómico, y en consonancia con la evolución mostrada por los datos agregados, se constata el avance de la productividad en manufacturas y el cambio en la composición del empleo por nivel de cualificación hacia niveles con mayor formación. Los resultados avalan la especial relevancia del capital humano, junto con la internacionalización y la efectividad de la inversión en I + D como determinantes de la productividad desde comienzos de los años noventa.

La innovación: ahora más que nunca

P. 139-153

Ascensión Barajas – Elena Huergo – Lourdes Moreno

Resumen

El objetivo de este trabajo es proporcionar evidencia empírica de las diferencias, si existen, en los resultados tecnológicos y económicos entre empresas manufactureras españolas con y sin perfil innovador, una vez que se controla por determinados efectos individuales predeterminados a la decisión de invertir en I + D + I. El análisis presta especial atención a los posibles cambios de comportamiento durante la presente crisis económica. Las estimaciones realizadas proporcionan evidencia de una mejor evolución global de las empresas con perfil innovador en términos de las innovaciones obtenidas, productividad, empleo o ventas. Este comportamiento más favorable se acentúa en el caso de las empresas que realizan actividades de I + D + I de forma persistente, incluso durante la crisis.

Contribución de las estrategias de internacionalización a los resultados de la empresa manufacturera española

P. 156-170

Raquel García-García – Esteban García-Canal – Mauro F. Guillén

Resumen

En el presente trabajo se analiza la expansión internacional reciente mediante Inversión Directa en el Exterior (IDE) de las empresas manufactureras españolas cotizadas y el impacto que esta ha tenido en su rentabilidad. Los resultados obtenidos indican que no hay una relación significativa entre el número de países hacia donde se expande la empresa manufacturera y su rentabilidad económica. No obstante, en el caso de la rentabilidad de mercado la relación sigue un patrón en forma de U invertida. Asimismo se observa que aquellas empresas que realizan un mayor porcentaje de sus inversiones en solitario generan mayores expectativas de creación de valor en el mercado de capitales.

La tecnología, la cualificación de la fuerza del trabajo y los sistemas productivos: impacto en la productividad en la empresa manufacturera española

P. 171-183

Alejandro Bello Pintado – Teresa García Marco

Resumen

Este artículo estudia el impacto de la tecnología y la cualificación de la fuerza del trabajo sobre la productividad utilizando una muestra de empresas de la industria española, analizando si este impacto depende de la configuración del sistema de producción mayoritariamente utilizado por las empresas y existen posibles complementariedades entre las nuevas tecnologías de fabricación y las habilidades de los trabajadores. Los resultados indican que tanto la utilización de tecnologías avanzadas de fabricación como la cualificación de la fuerza del trabajo tienen un efecto significativo sobre la productividad. Asimismo, se detectan efectos sinérgicos significativos dependiendo del sistema productivo

utilizado por la empresa.

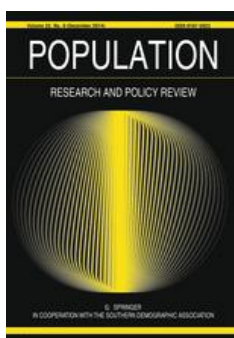
Los clústeres en España: ¿plancas para especialización inteligente?

P. 184-196

María José Aranguren – Susana Franco – Asier Murciego – James R. Wilson

Resumen

en este artículo se aborda un mapeo de los clústeres para las comunidades autónomas españolas con objeto de apoyar el desarrollo de las estrategias de especialización inteligente (S3) y la reindustrialización que promuevan la competitividad. Para ello, se presenta una tipología según el comportamiento exportador que sirve para identificar los clústeres que son más significativos en cada comunidad autónoma dada su relevancia, competitividad y dinamismo. Este ejercicio podría ser de utilidad para identificar las fortalezas territoriales y ayudar tanto en la selección de prioridades de las S3 de cada región como en la identificación de posibles sinergias entre comunidades autónomas.



Population research and policy review, ISSN 0167-5923
Volume 34, number 3 (june 2015)

The Impact of Social Security on Return Migration Among Latin American Elderly in the US

P. 307-330

Alma Vega

Abstract

International migration has long been considered the preserve of working-age adults. However, the rapid diversification of the elderly population calls for increased attention to the migration patterns of this group and its possible motivations. This study examines whether Latin American immigrants who are primary Social Security beneficiaries are more likely to return to their home countries during later life if they receive lower Social Security benefits. Using a regression discontinuity approach on restricted data from the US Social Security Administration ($N=1,515$), this study presents the results of a natural experiment whereby the Social Security Administration unexpectedly lowered the Social Security benefits of the 1917–1921 birth cohorts due to a miscalculation in the benefit-calculation formula. Results suggest that approximately 10% of primary Social Security beneficiaries from Latin America born close to these dates return migrated, the probability of which was not affected by Social Security benefit levels.

Ethnicity and Labor Market Incorporation of Post-1990 Immigrants in Israel

P. 331-359

Moshe Semyonov, Rebeca Rajjman, Dina Maskileyson

Abstract

Using data from “The Immigrants Survey” we compare economic incorporation of four ethnic groups of immigrants who arrived to Israel between 1990 and 2007: Ethiopia, Western Europe and the Americas, Asia and North Africa, and the Former Soviet Union. Labor market incorporation is evaluated in terms of labor force participation, occupational attainment and earnings. The analysis reveals that regardless of ethnicity, when compared to native-born, immigrant women face greater disadvantages in the labor market than immigrant men. Further analysis reveals that immigrants from the Former Soviet Union are more likely to become economically active than the other groups; immigrants from Europe and the Americas have better access to high status occupations than do either immigrant Former Soviet Union or Asia and Africa and Ethiopia. Ethiopian immigrants are the most disadvantaged group in attainment of high status lucrative occupations and earnings. The findings point toward an ethnic hierarchy among post-1990 immigrants in Israel with European-Americans at the top, followed by Soviet immigrants, followed by immigrants from Asia–Africa and ending with Ethiopian immigrants at the bottom. The meaning of these findings for possibility of emergence of a more diversified and elaborated system of ethnic stratification is discussed in light of Israel’s immigration policy.

Short-Term Labor Migration from Rural North India: Evidence from New Survey

P. 361-380

Data

Diane Coffey, John Papp, Dean Spears

Abstract

Despite high rates of internal migration, India is urbanizing relatively slowly. This paper uses new data from rural north India to study short-term migration to urban areas and its role in rural livelihoods. First, we demonstrate the importance of data collection techniques tailored to understanding short-term migration. Second, we consider how traditional theories of

migration apply in this context, where the fixed costs of migration are low, the opportunity costs vary by season, and where migration is negatively selective for education and economic status. We conclude by considering the implications of this migration for theories of development and development policies.

India's Age Structure Transition, Sectoral Labor Productivities, and Economic Growth: Evidence and Implications Based on National Transfer Accounts

P. 381-415

M. R. Narayana

Abstract

Using the new methodology of National Transfer Accounts, this paper quantifies the economic impacts of age structure transition and productivity growth rate on India's economic growth over the period 2005–2050 by formal and informal sectors. Growth effects are captured by the first demographic dividend (FDD) and distinguished by sector-specific (a) productivity age profiles, (b) relative and absolute labor productivity growth rates, and (c) population distribution for the benchmark year during 2004–2005. Empirical results show that in the presence of these sector-specific differences, growth effects are higher and the sources of lower and slower FDD are attributable to lower productivity levels, growth rates of productivity, and growth rate of effective number of producers in informal sector. Further, throughout, growth effects of productivity are found to be stronger than the age structure transition. Sensitivity results show that growth effects can be remarkably higher at an annual rate of 17% if benchmark output can be doubled in the informal sector, or FDD can be sustained up to 2050 if India's productivity profile in formal (and informal) sector has a comparable shape with that of Japan/USA (and Philippines/Indonesia/Nigeria). Overall implications show that stronger policy efforts are required for improvement in productivity levels and growth in informal sector to maximize long-run economic growth through FDD. These new results and implications may be of relevance for formulation of age-structure and informal sector related growth promotion policies in other developing countries of Asia, Latin America and Africa

Educational Attainment and Timing to First Union Across Three Generations of Mexican Women

P. 417-435

Rhiannon A. Kroeger, Reanne Frank...

Abstract

We use data from Wave 3 of the Mexican Family Life Survey ($N = 7,276$) and discrete-time regression analyses to evaluate changes in the association between educational attainment and timing to first union across three generations of women in Mexico, including a mature cohort (born between 1930 and 1949), a middle cohort (born between 1950 and 1969), and a young cohort (born between 1970 and 1979). Mirroring prior research, we find a curvilinear pattern between educational attainment and timing to first union for women born between 1930 and 1969, such that once we account for the delaying effect of school enrollment, those with the lowest (0–5 years) and highest levels of education (13+ years) are characterized by the earliest transition to a first union. For women born between 1970 and 1979, however, we find that the relationship between educational attainment and timing to first union has changed. In contrast to their peers born in earlier cohorts, highly educated women in Mexico are now postponing first union formation relative to the least educated. We draw on competing theories of educational attainment and timing to first union to help clarify these patterns in the context of Mexico

Exploring the Inequality-Mortality Relationship in the US with Bayesian Spatial Modeling

P. 437-460

Tse-Chuan Yang, Leif Jensen

Abstract

While there is evidence to suggest that socioeconomic inequality within places is associated with mortality rates among people living within them, the empirical connection between the two remains unsettled as potential confounders associated with racial and social structure are overlooked. This study seeks to test this relationship, to determine whether it is due to differential levels of deprivation and social capital, and does so with intrinsically conditional autoregressive Bayesian spatial modeling that effectively addresses the bias introduced by spatial dependence. We find that deprivation and social capital

partly but do not completely account for why inequality is positively associated with mortality and that spatial modeling generates more accurate predictions than does the traditional approach. We advance the literature by unveiling the intervening roles of social capital and deprivation in the inequality-mortality relationship and offering new evidence that inequality matters in US county mortality rates.



Statistics & risk modeling with applications in finance and insurance, ISSN 2193-1402
Volume 32, issue 1 (april 2015)

Moment based estimation of supOU processes and a related stochastic volatility model

P. 1

Stelzer, Robert - Tosstorff, Thomas - Wittlinger, Marc

Abstract

After a quick review of superpositions of OU (supOU) processes, integrated supOU processes and the supOU stochastic volatility model we estimate these processes by using the generalized method of moments (GMM). We show that the GMM approach yields consistent estimators and that it works very well in practice. Moreover, we discuss the influence of long memory effects.

Quasi-Hadamard differentiability of general risk functionals and its application

P. 25

Krätschmer, Volker - Schied, Alexander - Zähle, Henryk

Abstract

We apply a suitable modification of the functional delta method to statistical functionals that arise from law-invariant coherent risk measures. To this end we establish differentiability of the statistical functional in a relaxed Hadamard sense, namely with respect to a suitably chosen norm and in the directions of a specifically chosen “tangent space”. We show that this notion of quasi-Hadamard differentiability yields both strong laws and limit theorems for the asymptotic distribution of the plug-in estimators. Our results can be regarded as a contribution to the statistics and numerics of risk measurement and as a case study for possible refinements of the functional delta method through fine-tuning the underlying notion of differentiability.

Series expansions for convolutions of Pareto distributions

P. 49

Nguyen, Quang Huy - Robert, Christian Y.

Abstract

Asymptotic expansions for the tails of sums of random variables with regularly varying tails are mainly derived in the case of identically distributed random variables or in the case of random variables with the same tail index. Moreover, the higher-order terms are often given under the condition of existence of a moment of the distribution. In this paper, we obtain infinite series expansions for convolutions of Pareto distributions with non-integer tail indices. The Pareto random variables may have different tail indices and different scale parameters. We naturally find the same constants for the first terms as given in the previous asymptotic expansions in the case of identically distributed random variables, but we are now able to give the next additional terms. Since our series expansion is not asymptotic, it may be also used to compute the values of quantiles of the distribution of the sum as well as other risk measures such as the Tail Value at Risk. Examples of values are provided for the sum of at least five Pareto random variables and are compared to those determined via previous asymptotic expansions or via simulations.

Abstract

We propose a model for the computation of the loss probability distribution allowing to take into account the not-exchangeable behavior of a portfolio clustered into several classes of homogeneous loans. These classes are classified as 'large' or 'small' depending on their cardinality. The hierarchical hybrid copula-based model (HHC for short) follows the idea of the clusterized homogeneous copula-based approach (CHC) and its limiting version or the limiting clusterized copula-based model (LCC) proposed in our earlier work. This model allows us to recover a possible risk hierarchy. We suggest an algorithm to compute the HHC loss distribution and we compare this cdf with that computed through the CHC and LCC approaches (in the Gaussian and Archimedean limit) and also with the pure limiting approaches which are commonly used for high-dimensional problems. We study the scalability of the algorithm.
