

International  
Statistical  
Institute

Search

MEMBERSHIP

COMMUNITY

ABOUT US

CAPACITY

NEWS

EVENTS

PUBLICATIONS

RESOURCES

---

# ISI WSC 2021 Proceedings

In this section:

- [Events](#)
- [World Statistics Congresses »](#)
- [Other conferences](#)
- [Sponsorship or endorsement](#)
- [Webinars](#)
- [Courses](#)



186 results

## Day 1

*IPS 029:* Food security in Africa? Measurement approaches and recent trend

[Measuring food security, challenges of statistical systems –](#)

Said Saghir Zarouali

*IPS 086:* The statistics of extremes

[Coupling clustering algorithms and extreme value theory to define spatial](#)

[homogenous hydrological regions](#) –

Philippe Naveau

*IPS 119*: Digital Data: what innovations in quality management are needed to keep the momentum to 2030?

[Error Analysis for Hybrid Estimates that use Big Data](#) –

Dennis Trewin

*IPS 183*: Methodological advancements in nowcasting development indicators

[Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks \(LSTM\)](#) –

Daniel Hopp

[Remote but well connected? Neighboring but isolated? The measurement of remoteness in the context of SIDS](#) –

Fernando Cantu

[The Eurostat business cycle clock and the pandemics: some preliminary considerations](#) –

Rosa Ruggeri Cannata

## *CPS 023*: Statistical analysis for high-dimensional data

[Big Data Mining from Big Data Sources](#) —

Md Shariful Islam

[Convergence rates of model-assisted estimators in high-dimensional settings](#) —

Mehdi Dagdoug, Camelia Goga, David Haziza

[High-Dimensional Temporal Disaggregation](#) —

Luke Mosley, Idris Eckley, Alex Gibberd

## *CPS 034*: Topics in Business and Industry

[Achieving due diligence in the supply chain - how statistics can help](#) —

Oscar Lemmers

[Fast Product Optimization with Uniform Design, Application in Cosmetic](#) —

Lina Zhao

## *CPS 031*: Statistics in Practice-Some Issues III

A multinomial multilevel  
model for measuring  
continuity and regularity of the  
Italian museums opening –

Sandra De Iaco, Sabrina  
Maggio, Claudia Cappello

Discriminating between and  
within (semi)continuous  
classes of both Tweedie and  
geometric Tweedie models –

Rahma Abid, Célestin C  
Kokonendji

## *CPS 018: Pandemic I*

Impact of Covid-19 on the  
structure of GDP by type of  
economic activity (Ukrainian  
case) –

Ruslan Motoryn, Tetiana  
Motoryna, Kateryna  
Prykhodko

Mapping the Potential Use of  
ICT for Distance Learning  
During Covid-19: Demand and  
Supply-Side Approach –

Taly Purwa, Diyang Gita  
Cendekia

Statistical modelling of  
covid-19 deaths in Nigeria –

George Uchechukwu, Okereke  
Emmanuel Wilfred

The Impact of Covid-19

Pandemic on Tourism in the  
GCC –

Amira Al-Salhi

The Impact of Movement  
Restrictions during Covid 19  
on Concentrations of Nitrogen  
Dioxide in GCC Countries –

Ibtihaj Alsiyabi

## *CPS 017: Official Statistics I*

Modeling international  
migration flows by integrating  
multiple data sources –

Emanuele Del Fava, Arkadiusz  
Wisniowski, Emilio Zagheni

The use of administrative tax  
data as an estimation strategy  
–

Sihle Khanyile

## *CPS 012: Global Economics & Development III*

Integrated Real Time Tourism  
Data for Optimizing  
Indonesian Economic Growth  
from Tourism Sector (Case  
Study in 5 Super Priority  
Tourism Object in Indonesia)  
–

Christiayu Natalia, FX Gugus  
Febri Putranto

Research on Promoting  
Growth and Benefiting  
People's Livelihood Effect of  
Rural Dilapidated Housing  
Reform —

Du Zhixiu, Du Jinzhu

The State of Migration and  
Mobility Data in Europe : a  
Systematic Assessment —

M. J. Danko, E. Del Fava, D.  
Jasilionis, D. A. Jdanov, E.  
Zagheni

## *CPS 009*: Global Economic and Social Questions II

Statistical approach to the  
impact of air pollution on the  
otolaryngology system  
diseases —

Dominik Nowakowski, Barbara  
Jasiulis-Goldyn

## *CPS 001*: Analyzing Spatial Data

Analysis of Influencing  
Factors on Population Change  
in Honam Region Considering  
Spatial Autocorrelation —

Kiseok Song

Measuring spatial distribution  
of education inequality —

[of education inequality](#)

Siti Nurliza Samsudin

## *IPS 132: Web data for official statistics: the ESS web intelligence hub*

[Big data, better skills intelligence? : Using online job advertisements to map skills trends](#) —

Jasper van Loo

[The Web Intelligence Hub – A tool for integrating web data in Official Statistics](#) —

Albrecht Wirthmann

## *IPS 157: Poverty and Inequality: New Challenges and New Statistical Responses*

[Poverty and Inequality in Germany, Conclusions Derived from a Hegelian Overall Picture of Economic Activities in Human Society and Nature](#) —

Helmut Maier

[Poverty Measuring: New Challenges and Statistical Response](#) —

Irina Eliseeva , Maria Dekina

[Risk estimation of COVID-19](#)



RISK ESTIMATION OF COVID-19  
spread between poverty  
clusters in Africa using  
Markov Chain modeling –

Mintodê Nicodème Atchadé ,  
Gandonou Sèdjro Amos ,  
Svetlana Vladimirovna  
Kurisheva

## *IPS 174: Central Banks as producers of official statistics*

Remittances and their impact  
of poverty: the case of Albania  
–

Elona Dushku, Argita Frasheri

The more the merrier:  
enhancing traditional cross-  
border portfolio investment  
statistics using security-by-  
security information –

Antonio Rodríguez Caloca

Unbundling Package Tours: a  
Machine Learning Application  
with the LASSO –

Andrea Carboni

## *IPS 216: Fintech Statistics*

The Use of AI for Data  
Gathering - Finding and  
Monitoring Fintechs –

Guillaume Belly , Andy Bosyi ,  
Elizabeth D...

Elisabeth Devys , UIT von  
Kalckreuth

*IPS 162: Innovation on  
Energy Statistics and  
beyond to meet  
environmental targets  
and create a carbon  
neutral economy*

[Developing a Methodology for  
Measuring Fossil Fuel  
Subsidies –](#)

Clare O'Hara

[Estimated CO2 emissions  
from energy use in the EU –](#)

Stavros Lazarou

---

## Day 2

*IPS 077: Corner Pillars of  
Probability Education -  
Part I*

[Early schooling to shape the  
basic elements of probabilistic  
literacy –](#)

Laura Martignon

[Why is it necessary to teach  
probability from kindergarten](#)

onwards? –

Teresita Evelina Terán

## *IPS 215: Statistical literacy, visualization and communication*

When reaching is no longer enough: 8 tips to engage with Central Banks' data users –

Lígia Nunes

## *IPS 181: The Use of Big Data to Predict Economic Activity During Covid-19 Pandemic*

The Use of Mobile Positioning Data as Proxy of Economic Activity in Indonesia –

Amanda Pratama Putra

## *IPS 192: Population Size Estimation*

Bayesian analysis of one-inflated models for population size estimation –

Titiana Tuoto

Multiple record system estimation with overlapping lists, where some lists cover only part of the population –

Peter van der Heijden

*SIPS 237*: ISI Jan  
Tinbergen Awards  
Session

[High-dimensional generalized  
semiparametric model for  
longitudinal data](#) –

Mozhgan Taavoni

*IPS 047*: Privacy  
Preserving Computation  
Techniques in Official  
Statistics

[Input Privacy: Towards a  
Logical Framework for  
Defining Official Statistics  
Scenarios](#) –

Monica Scannapieco

[Key GDPR Legal Issues  
regarding Privacy-Preserving  
Techniques](#) –

Kuan Hon

*IPS 126*: Labeling Issues  
in Machine Learning  
Methodology

[Impact of Mislabelling on  
Deep Learning Methods and  
Strategies for Improvement](#) –

Marcel Dettling , Martin Frey ,  
Manuel Walser , Patrick Haas

*IPS 024: Methodological responses to the COVID-19 pandemic by National Statistical Institutes*

[Ensuring statistical production and quality in the COVID-19 pandemic: the Italian experience –](#)

Stefano Falorsi , Orietta Luzi ,  
Monica Scannapieco , Mauro Scanu

[The Experience of Stats NZ in being part of the New Zealand agency managing the COVID-19 Pandemic –](#)

Vince Galvin

*SIPS 222: Telling Truth with Scientific Data*

[Data on Social and Environmental Determinants of Health and COVID-19: Public Use and Sharing –](#)

Kevin Lane

*IPS 135: Seasonal adjustment of official statistics disrupted by COVID-19 effects*

[Treatment of Time Series](#)

## Outliers via Maximum Entropy

—

Tucker McElroy

## *IPS 042*: Counting People with Administrative Data instead of a Traditional Census

The use of Administrative Data in the production of Italian Permanent Census Estimates —

Nicoletta Cibella

## *CPS 036*: Topics in Environmental Statistics

Chimney fire prediction based on environmental variables —

Changqing Lu, Marie-Colette Van Lieshout, Maurits De Graaf, Paul Visscher

Groundwater time series analysis with clustering and LSTMs for sporadic monitoring datasets —

Stephanie Clark

## *CPS 027*: Statistical Modeling III

Supervised Text Classification with Leveled Homomorphic Encryption —

Zachary Zanussi, Benjamin Santos, Saeid Molladavoudi

[Universal inference with composite likelihoods](#) —

Hien D. Nguyen, Jessica Bagnall-Guerreiro, Andrew T. Jones

## *CPS 020*: Probability and Statistical Models

[On the study of extinction of circuit chains describing random walks in random environments](#) —

Chrysoula Ganatsiou

[Optimal covariance matrix estimators in a seemingly unrelated regression model](#) —

Shun Matsuura, Hiroshi Kurata

[Regularized Double Machine Learning in Partially Linear Models with Unobserved Confounding](#) —

Corinne Emmenegger, Peter Bühlmann

[Statistical analysis of discrete-valued time series based on parsimonious models](#) —

Yuriy Kharin

## *CPS 011*: Global

## Economics & Development II

Access to drinking water:  
complementary indicators to  
monitoring goal 6.1 of 2030  
Agenda at local level in Brazil  
—

Ana Clara Cancelo Cruz,  
Andrea Diniz da Silva

Implementation of  
Marketplace Data in The  
Calculation of The Consumer  
Price Index (Case Study of CPI  
Cities in Java, Indonesia) —

Muhammad Ghazy Al  
Haqqoni, Setia Pramana

Linking census of agriculture  
and business register in Brazil:  
some developments and  
many challenges —

Andrea Diniz Da Silva, Raphael  
Molina Guimaraes, Geremias  
De Mattos Fontes Neto

Monitoring the effects of  
drought, poverty and access  
to food in low-income  
communities of the city of  
Manzini, Eswatini —

Sipho Felix Mamba

The Impact of Mining Sector  
on Economic Growth and  
Unemployment Rate in East  
Kalimantan, Indonesia: A  
Simultaneous Panel Data



[Analysis using EC2SLS](#) –

Akhmad Fatikhurriqzi, Robert Kurniawan

*CPS 019: Pandemic II*

[Comparing the safety of COVID-19 vaccines: a geometrical approach](#) –

Silvia Columbu, Beniamino Cappelletti-Montano, Stefano Montaldo, Musio Monica

[Compiling Quarterly GDP by Economic Activity in Volume Measures during COVID-19 Pandemic](#) –

Alex Li Sze-chun, Howard Wong Ho-fai

[Covid-19 And Its Impact In The Manufacturing Sector : A Case Study In the State of Perak, Malaysia](#) –

Maslina Binti Samsudin, Nur Aizatul Hajar Binti Abdul Kadar

*CPS 005: Environmental Analysis*

[A model for spatial binary data](#) –

Gabrielle. E. Kelly, J. Pan

[A three-way synthetic air quality index capable of assessing health risks](#) –

Passamani Giuliana, Masotti  
Paola

[Determinants of Improved  
Drinking Water Access in  
Bengkulu and Papua](#) –

Hanin Rahma Septina, Aulia  
Dini

[Modeling the impact of air  
pollution on the respiratory  
system diseases](#) –

Barbara Jasiulis-Goldyn,  
Bozena Cegiela

## *CPS 033*: Temporal Modeling

[Granger Causality Analysis for  
Functional Time Series Data](#) –

Shubhangi Sikaria, Rituparna  
Sen

## *CPS 035*: Topics in Computational Statistics

[Analysis of daily precipitations  
from synoptic stations using  
non-homogeneous hidden  
Markov model](#) –

N. Bokossa, V. Adanhounme

[Baseline Methods for  
Bayesian Inference in Gumbel  
copula](#) –

Mario M. Pizarro, Jacinto  
Martín Eva I. Sanjuán María I.

Parra, Eva L. Sanjuan, Mariana L.  
Parra

*IPS 131: Reformed  
Journal of Data Science  
in the Era of Data  
Science*

[Sparse Learning with Non-convex Penalty in Multi-classification](#) –

Nan Li

*IPS 137: Meta-analysis  
Methods with  
Applications in Health  
Sciences*

[Bayesian hierarchical analysis of equity in delivery care services coverage for adolescents in 54 low-and middle-income countries, 2000-2030](#) –

Mizanur Rahman

*IPS 187: National  
ownership to SDG  
monitoring -  
Achievements and  
Challenges*

[Engaging with stakeholders and responding to policy priorities as part of the 2021 VNR process, Lao PDR](#) –

Phetsamone Sone

## *IPS 027: Advances in Sports Statistics (2 of 2)*

[Using Higher-Order PLS-PM model for performance analysis in football –](#)

Mattia Cefis

## *IPS 212: Relevance of data integration and advanced analytics in central bank decision-making*

[The future of data collection & data management: Agile RegOps for digitalizing the regulatory value chain –](#)

Daniel Muench

---

## Day 3

## *IPS 050: Measuring SDG Progress*

[A note on assessing progress towards the SDG targets –](#)

Marcello D'Orazio

## *IPS 043: Trusted Smart Surveys*

Trusted Smart Surveys:  
European Platform's  
Methodological Framework –

Claudia De Vitiis

Trusted Smart Surveys:  
Solutions for the European  
Statistical system - An  
overview of the objectives and  
the main challenges –

Shari Stehrenberg

## *SIPS 217: Unintended Consequences of COVID-policy on Health and the Environment*

Public health response to the  
COVID-19 pandemic: the  
forecaster's dilemma and  
long-term consequences –

Mika Naumanen

## *IPS 141: IASS Tribute Session to Lars Lyberg: How survey experts can contribute to the better use of citizen-generated data for public policy?*

Community led monitoring  
accelerating pace of  
empowerment and

Empowerment and  
accountability –

Shikha Shrestha

*IPS 100*: On the  
visualisation of digital  
socio-economic data

The Cressie-Read Divergence  
Statistic and Digital Data  
Visualisation –

Eric J. Beh, Rosaria Lombardo

*SIPS 184*: Effectiveness  
of the outreach of official  
statistics standards and  
guidelines,  
methodologies and  
recommendations to  
developing statistical  
systems

Standards in Social statistics:  
the capacity of national  
statistical systems to adapt to  
the international standards,  
problems, challenges –

Lisa Grace S. Bersales

*CPS 021*: Quality of Life

Outlier detection for  
compositional data by  
municipality and disabled  
people in Brazil –

Paulo Tadeu Meira e Silva De

Oliveira

## *CPS 028*: Statistics in Education

[A New Statistical Adventure  
with Exploristica 2.0](#) —

Martins J. Pinto, Campos  
Pedro

[Statistical Literacy  
Competencies as Post-  
modern Basic Civic Skills](#) —

Jukka Hoffrén

## *CPS 006*: Estimation and Regression

[Estimation and model  
selection for mixing graphical  
models](#) —

Magno T.F. Severino, Florencia  
Leonardi

## *CPS 003*: Bayesian Modeling

[A hierarchical  
multidimensional graded  
response model with  
application to the study of  
sustainability perception](#) —

Marcelo Andrade Da Silva,  
Jorge Luis Bazán, Edna  
Possan, Silvana Ligia Vincenzi

## *CPS 032: Statistics in Practice-Some Issues IV*

A study on Transition from Labour to Capital Intensive Industry in Electrical, electronic and optical products activity –

Rabi'atul'adawiah Shabli

Gold Investment in Indonesia: Prediction and Prospective in Pandemic Era Using LSTM Deep Learning –

Atika Kautsar Ilafi, Lita Jowanti, Fadhilah Annisa Nur

Informality and urban planning in the global South: The case of Masvingo city, Zimbabwe –

Average Chigwenya

## *CPS 013: Health and Biomedical Analysis I*

A comparison of sequential design procedures for discriminating enzyme kinetic models –

Elham Yousefi

## *CPS 025: Statistical Modeling I*

Bayesian selector of adaptive bandwidth for multivariate



gamma kernel estimator with  
standard bias correction on  $[0, \infty)^d$  –

Somé Sobom M, Kokonendji  
Célestin C., Adjabi Smail,  
Beddek Said

Profile Transformation for  
Reciprocal Averaging and  
Singular Value Decomposition  
–

Ting-Wu Wang, Eric J. Beh

Stochastic block model for  
multiple networks –

Tabea Rebařka

*IPS 166*: Economic  
activities classifications -  
review and  
implementation

Capacity building for accurate  
economic classifications in a  
rapidly changing world: a  
European case –

Mari Rantanen

Metadata Modelling for  
Economic Statistical  
Classifications –

Andrew Hancock

*IPS 106*: Exploring  
Statistics on Tourism,  
Migration, Population

## Migration, Population and Displacements by Using Mobile Positioning Data

Displacement and Disaster  
Indicators Derived from  
Mobile Positioning Data –

Ayumi Arai

Mobility data brings new  
insights to local cultural  
tourism management: Case of  
Setomaa in South Estonia –

Siim Esko

What are similarities and  
differences in deriving  
migration and tourism  
statistics from mobile  
positioning data? –

Shorena Tsiklauri, Mariana  
Jalagonia, Titi Kanti Lestari

## *IPS 214*: Addressing globalisation challenges

A Typology of Captive  
Financial Institutions and  
Money Lenders (sector S127)  
in Luxembourg –

Gabriele Di Filippo

An assessment of euro area  
households' missing foreign  
assets –

Martin Schmitz

## *IPS 018: Census of agriculture and COVID-19: impact and lessons*

Running Portuguese  
Agricultural Census in  
Pandemic Crisis –

Carlos Coimbra

## *IPS 148: Statistical Literacy Across Disciplines*

Control Charts and Capability  
Analysis for Statistical  
Process Control –

Arun Kumar Sinha

Uncertainty evaluation with  
the concepts of Bayesian  
Statistics –

Richa Vatsa

## *IPS 090: A new survey model to bridge the agricultural data gap*

Integrated sampling design  
for agricultural and socio-  
economic households  
surveys: a cost effective  
approach for agricultural and  
rural statistics –

Dramane Bako

## *IPS 078: Corner Pillars of Probability Education - Part II*

A complex concept about statistical inference and a planned school experiment based on it –

Ödön Vancsó, Manfred Borovcnik, Peter Fejes-Tóth

Corner pillars of probability literacy –

Manfred Borovcnik

Reasoning with Risk: Teaching probability in the time after COVID –

Ramesh Kapadia

---

## Day 4

*IPS 039: Time series forecasting: Computational challenges and application to electricity load forecasting*

Generating experts by boosting diversity –

Jean-Michel Poqai

~ ~

## *IPS 244: Developing Machine Learning Methods in Experimental Official and Analytical Statistics*

[Machine Learning Methods  
for Processing and Analyzing  
Integrated Datasets of Official  
Statistics](#) —

Elena Zarova

## *IPS 175: New developments in central bank statistics around the world*

[Nowcasting and monitoring  
Israeli real economic activity](#)  
—

Tim Ginker

[Revealing investors' sentiment  
amid COVID-19: the Big Data  
evidence based on internet  
searches](#) —

Jean Christine A. Armas,  
Pamela Kaye A. Tuazon

## *IPS 233: Data Science & Statistics*

[Variable screening in high  
dimensional regression via](#)

[random projection ensembles](#)

—

Angela Montanari

*IPS 180*: Lessons learned  
for macroeconomic  
statistics from the  
Covid-19 pandemic

[Macroeconomic statistics  
during the Covid-19 pandemic  
— a user perspective —](#)

Christoph Maier

*CPS 004*:  
Business/Manufacturing

[Modified Weakest-Link Load-  
Sharing Model and Its  
Application to the Breakdown  
Process of High-k Gate  
Dielectrics —](#)

David Han

*CPS 014*: Health and  
Biomedical Analysis II

[Bayesian modeling of health  
state preferences: could  
borrowing strength from  
existing countries' valuations  
produce better estimates —](#)

Samer A. Kharroubi

[On nonparametric Phase I  
analysis of individual](#)

observations from a change-  
point model perspective —

Christina Parpoula

## *CPS 015: Models in Official Statistics I*

Development of optimal  
weights for cyclical composite  
indicator systems based on  
multivariate spectral analysis  
of time series —

Víctor Alfredo Bustos y de la  
Tijera, Noemí Lopez

The use of scanner data to  
compile regional price indices  
in Germany with an outlook to  
income inequality —

Chris-Gabriel Islam, Alina  
Happ

## *CPS 024: Statistical Literacy*

Dropout and completion in the  
Bachelor in Statistics in Brazil:  
use of survival analyses in a  
competing risk scenario —

Adriane Caroline Teixeira  
Portela, Andrea Diniz da Silva,  
Giovana Oliveira Silva

Ranking of universities in  
Statistics: a shift towards  
excellence? —

Veljko Jeremic, Srda  
Bieladinovic, Mladen

DJERADUHOVIC, MIRAUER  
Stamenkovic

## *CPS 026: Statistical Modeling II*

[Constructing a Realized DCC  
Model with Measurement  
Errors](#) —

Hideto Shigemoto, Takayuki  
Morimoto

[D-Optimal designs for  
Antoine's Equation:  
with homoscedastic and  
heteroscedastic response](#) —

Carlos De la Calle-Arroyo,  
Jesus Lopez-Fidalgo, Licesio  
J. Rodriguez- Aragon

[Sequential criticality test for  
branching process with  
immigration](#) —

Junfan Tao, Keiji Nagai,  
Kotaro Hitomi, Yoshihiko  
Nishiyama

## *CPS 029: Statistics in Practice-Some Issues I*

[Development as a Matter of  
Statistics: How an Integrated  
Statistical System Can  
Contribute to the  
Development of a Country](#) —

Seyyed Mahdi Sadati  
Nooshabadi



[Integrating Quality  
Management and Enterprise  
Architecture in Hungary](#) –

Andrea Petres, Józsefné  
Ferencz, Zoltán Vereczkei

[Retreat towards systemic  
reform. What stakes?](#) –

Nisrine Ghefou

*CPS 037: Topics in  
survey sampling*

[Comparison of Selection  
Methods for Lattice Sampling](#)  
–

João Gabriel Malaguti

*IPS 177: Communication  
and dissemination in the  
ESS in times of Covid-19*

[Istat communication strategy  
in times of pandemic crisis](#) –

Giulia Mottura

*IPS 213: Implications of  
the Digital Economy for  
Official Statistics*

[Effective measurement of the  
economy in the emerging  
digital age](#) –

Antonio Colangelo, Francis  
Gross, Florian Schuster

New Hedonic Quality  
Adjustment Method Using  
Sparse Estimation –

Sahoko Furuta

## *IPS 179*: Measuring illicit financial flows in official statistics

Can we measure illicit  
financial flows? Methods for  
the measurement of tax and  
commercial illicit financial  
flows –

Bojan Nastav

The importance of data and  
statistics in the field of illicit  
financial flows (IFFs) –

Bishop Akolgo

What are illicit financial flows?  
Concepts for their  
measurement in official  
statistics –

Diana Camerini

## *IPS 207*: Statistical Modeling of Simulated Gaming Methodology

Statistical Modeling of  
Simulated Gaming  
Methodology –

Siddhant Karmali

*IPS 186: Strengthening  
Administrative data:  
what countries have  
learned and applied from  
the administrative data  
collaborative*

[Strengthening Administrative  
data: what countries have  
learned and applied from the  
administrative data  
collaborative –](#)

Jorge Urrutia

[Strengthening Administrative  
Data: what countries have  
learned and applied from the  
administrative data  
collaborative –](#)

Phub Sangay

*IPS 088: The Mirror has  
Two Faces: An Analysis  
of Asymmetries in  
Official International  
Merchandise Trade  
Statistics*

[Analysis of Asymmetries in  
Official International  
Merchandise Trade Statistics:  
A Case Study of Malaysia –](#)

Siti Haslinda Mohd Din

[Bilateral Trade Asvmmetries in](#)

[International Policy Symposium on  
Asia Pacific](#) —

Janine De Vera

---

## Day 5

*IPS 011: Creating a  
better world using  
composite indicators*

[Cost and performance: a  
composite indicator for  
separated waste collection in  
Italy](#) —

Carlo Cavicchia , Pasquale  
Sarnacchiaro , Maurizio Vichi

*IPS 153: Enhancing  
growth and productivity  
indicators based on  
European national  
accounts*

[Enhancing growth and  
productivity indicators based  
on European national  
accounts: Selected capital  
productivity measures based  
on EU National Accounts data](#)  
—

Robert Stehrer

*IPS 105: HOW to fund  
official statistics  
production to support  
statistical principles and  
ethics*

[Some issues of funding of  
official statistics](#) –

Hallgrímur Snorrason

*IPS 087: Statistical and  
Data Literacy in Policy-  
Making*

[Statistical Literacy for Policy  
Makers](#) –

Milo Schield

[Towards a Framework for  
Data and AI Literacy](#) –

Katharina Schüller

*IPS 051: Alternative  
methods for SDG data  
disaggregation*

[The FAO Guidelines on data  
disaggregation for SDG  
Indicators using survey data](#) –

Piero Falorsi

*CPS 002: Applications in  
Banking and Finance*

[A comparative study between](#)

support vector machine and  
support vector data  
description in bank  
telemarketing –

Han Gao, Pei Shan Fam, Heng  
Chin Low

Research Study of Business  
Cycle and Early Warning  
Indicators for the Economy of  
Hong Kong – Challenges of  
Forecasting Work amid the  
COVID-19 Pandemic –

Eddie Kwok Ming-lok, Sharon  
Ng Pun-wai, Brian Cheng Chi-  
yan

Risk and Return of the  
Tontine: A Brief Discussion –

Peter Pflaumer

The Japanese Yen Interest  
Rate Swap Market in the Time  
of COVID-19 –

Shiori Inoue, Shota Miki,  
Yasufumi Gemma

## *CPS 007: Family and Gender*

Analysing Gender Equality at  
the Firm Level –

Dania Eugenidis

Developing a global living  
standards measure using  
census microdata –

Rodrigo Lovatón, Sula Sarkar

Reynolds Lovatari, Gaida Satriani

How Awareness Raising  
Program Affects the Bio-social  
Correlates of Adolescent's  
Consciousness on Sexual  
Harassment: A Multigroup  
Simultaneous Equation Model

—

Injamul Haq Methun

M-Quantile Small Area  
Estimation for Household Per  
Capita Income: A Case Study  
of Sub-Districts in Special  
Region Yogyakarta, Indonesia

—

Aldi Rochman Nulkarim, Ika  
Yuni Wulansari

*CPS 008*: Global  
Economic and Social  
Questions I

Official Business Sentiment  
Indicators for Tracking the  
Economy's Pulse and its  
Forewarning Capability —

Jason C.T. Chan, Amy S. S. Yu

*CPS 010*: Global  
Economics &  
Development I

Is Risk the Fuel of the  
Business Cycle? —

Christoph Schult

Statistical analysis of  
categorical data: the case  
study is based on the  
demographic and socio-  
economic determinants of  
access to clean and safe  
water among Basotho –

Moliehi Mokete

The Impact of Rural Inflation  
on Urban Inflation in Indonesia  
–

Masarina Flukeria, Mohamad  
Ikhsan

## *CPS 016*: Models in Official Statistics II

Fitting Cobb-Douglas function  
in modelling the impact of  
COVID-19 on the Malaysia's  
economy –

Muhammad Shafiq Harun

Three years of the crime and  
victims survey in New Zealand  
– lessons and perspectives –

Michael Slyuzberg, Tadhg Daly  
, Kate Preston , Sandra Geng ,  
Tianying Chu

Weighting a non-probability  
web sample survey carried  
out during the pandemic –

Marcelo Trindade Pitta, Pedro  
Luis Do Nascimento Silva



## *CPS 022: Sampling Design and Estimation*

[Design-based composite  
estimation rediscovered](#) —

Andrius Ciginas

[Integrated agriculture surveys:  
standards and country  
customization](#) —

Flavio Pinto Bolliger

[Perceptions about the  
impacts of covid-19 in human  
rights: preliminary results](#) —

Bernardo B. Martins, Andrea D.  
Silva, Elizabeth B. Hypolito,  
José Ricardo F. Cunha, Ana  
Clara C. Cruz, Bruno F. A. Paz,  
Fausto M. Pinheiro Jr.,  
Raphaella S. R. C. Lima

## *CPS 030: Statistics in Practice-Some Issues II*

[Disclosure risk assessment  
for frequency tables of a  
Brazilian economic survey](#) —

Sâmela Batista Arantes,  
Maysa Sacramento de  
Magalhães, José André de  
Moura Brito

[Measuring Environmental  
Health Inequality between the  
Rich and the Poor : Evidence  
from Indonesia](#) —

Kadek Swarniati, Ayesha  
Tantriana, Sri Hartini  
Rachmad

*IPS 067: Capturing  
Informal and Micro  
Economy in the Era of  
Digital*

[Malaysia's International Trade  
in Services: Unleash the cross  
border trade in digital services](#)

—

Kumutha Shanmugam

*IPS 173: Cooperation  
issues between NSOs  
and Central Banks*

[Recent Developments in the  
Use of Disclosed Detailed  
Financial Information in the  
Estimation of Japan's Flow of  
Funds Accounts](#) —

Yukiho Ishigami

[Uses of Mirror Data -  
Estimation of Foreign Assets  
and Liabilities of Households](#)

—

Swapan-Kumar Pradhan

*IPS 139: Data Science in  
Estimating Informal  
Sector and Gig Workers*

[Digitization and the informal economy: Scope and measurement challenges](#) –  
Jacques Charmes

[Estimating Informal Employment in Malaysia](#) –  
Hawati Abdul Hamid

*IPS 112: Live Cumulative Meta-Analysis for a Better World*

[The New World Order and the Challenges of Official Statistics in Nigeria](#) –  
Osuolale Peter Popoola

*IPS 185: Stochastic Processes and their Applications in Evolutionary Biology. New Trends*

[Branching random walks applied to modeling of germinal center reaction](#) –  
Daria Balashova

[Simulation of branching random walks in random media](#) –  
Vladimir Kutsenko

[Two-type Branching Random](#)

Walks with Different  
Configurations of Branching  
Sources —

Yulia Marakova

*IPS 019*: The New World  
Order and the Challenges  
of Official Statistics in  
Nigeria

Building Robust Collaboration  
between the Producers and  
the Users of the Official  
Statistics in Africa —

Nureni Olawale Adeboye

Official Statistics That Leave  
No one Behind: An  
Opportunity for Nigeria. —

Aliyu Usman

Revolutionizing Official  
Statistics in Nigeria —

Kelechi Acha

The Structure of National  
Bureau of Statistics and Data  
Procurement in Nigeria —

Iyabode Favour Oyenuga

---

*Day and session  
unknown*

Image Analysis of Tree-Rings  
Using Convolutional Neural  
Networks and the  
Classification of Infra-red  
Camera Images by CNN  
Analysis: A Study on Regional  
Revitalization by Improving  
the Reliability of Timber  
Traceability –

Miyoko Asano, Yoshiharu  
Asano, Hongjie Zheng

Soil organic carbon and the  
hydrology of rubber soil  
versus forest soil: a depth  
resolved linear regression  
analysis –

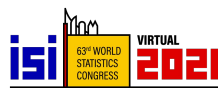
Indulekha Kavila, Bhava V Hari

Utilizing Nighttime-light (NTL)  
Data to Measure Indonesia's  
Gross Domestic Regional  
Product (GDRP) –

Agung Andiojaya, R.R.  
Nefriana, Riana Rizka, Titi  
Kanti Lestari

---

ISBN 978-907359239-1



Tiziana Tuoto

## Bayesian analysis of one-inflated models for population size estimation

Tiziana Tuoto<sup>1,2</sup>; Davide Di Cecco<sup>2</sup>; Andrea Tancredi<sup>2</sup>

<sup>1</sup> Italian National Institute of Statistics, Istat

<sup>2</sup> Sapienza University of Rome

### Abstract

The treatment of one inflation in estimating the size of an elusive population has received an increasing attention from researchers in recent years. The inflation of counts "one", that is the presence of an excess of "one"s, larger than compatible with any distributional assumption, deserves specific attention due to its effect on usual estimators for the population size. One-inflation may lead to overestimation of the usual estimators when not accounted for, while in general, when the observed counting model presents other sources of heterogeneity and this is ignored by the estimation procedure, it may produce severe underestimation in the population size. In this paper we propose a Bayesian approach that considers one inflation in count data models for capture recapture. An application to real data for the estimate of the size of some illegal populations is used to illustrate the proposed methodology.

### Keywords:

Capture-recapture; Illegal populations; Zero-truncated One-inflated Count Data Models; Bayesian Model Selection

### 1. Introduction

A popular methodology to estimate the size of hidden populations is the capture-recapture methods, originally used to estimate animal abundance. In capture-recapture, data can be collected at specific time points, therefore for each unit seen at least once there is a capture history. The data used in this work are a specific form of capture-recapture data. We collect data in continuous time hence we have only the total number of times a unit is captured. By means of an observational mechanism, i.e a register, a trapping system, a set of interviewers, a diagnostic devise, we can identify some population units and follow them in a specific time span. In this way, the observational mechanism provides us a list of population units with the count (i.e. the number of times) that they appear in the list. In the list we can observe individuals who are observed/captured 1, 2, 3, ..., times, however we cannot observe units not caught by the observational system. Hence, the list can be considered as incomplete and we want to estimate the hidden part of the population, i.e. the size of it not reported by our observational mechanism. Capture recapture data in this setting are usually called repeated counting data. To estimate the population size, one needs to model the counting process of observation/capturing.

In this framework, an increasing attention has been devoted to the inflation of counts "one", i.e. the presence of an excess of "one"s, larger than compatible with any distributional assumption, see Godwin and Böhning (2017), Godwin (2017), Godwin (2019), Böhning, Kaskasamkul, and van der Heijden, (2018), Böhning and van der Heijden (2019), Böhning and Friedl (2021).

One-inflation can occur for different reasons; for instance, we observe one-inflation when some units of the population cannot be captured anymore after the first capture. This may be the case of some wild animal populations, when the animals that experienced the capture once, find it so unpleasant that some of them develop the desire and/or ability to avoid subsequent captures. A similar reasoning can be applied also to human populations, particularly when the first capture consists of law enforcement, involves imprisonment or reveals an undesirable characteristic/behaviour. See Godwin and Böhning (2017) for a rich discussion on justifications and conditions for one-inflation in capture-recapture, also including an interpretation of one-inflation as limiting case of the so-called "trap shy" behavioural model.

One-inflation deserves specific attention due to its effect on usual estimators for population size. In general, when the observed counting model presents heterogeneity and this is ignored by the estimation procedure, it may produce severe underestimation in the population size. On the contrary, one-inflation may lead to overestimation of the usual estimators when not accounted for. This is true even for the well-known lower bound Chao estimator, which may severely overestimate the population size in the presence of one-inflation, as discussed in Chiu and Chao (2016) and in Böhning, Kaskasamkul, and van der Heijden, (2018).

In this paper we propose a Bayesian approach to count data models with one-inflation, for use in Horvitz–Thompson estimator of the population size. Under given conditions, the Bayesian approach results in population size estimates similar to the maximum likelihood ones, with the advantage of producing the credible intervals of the estimates as a by-product of the estimation procedure. Another advantage of the Bayesian approach is to allow incorporating previous knowledge on the hidden size of the population in the analysis, in the very favourable case in which this information is available to the analyst. Moreover, the Bayesian analysis allows testing the one-inflation assumption in a very natural way.

We apply our Bayesian proposals to real data for estimating the size of some illegal population, using also some popular dataset available from the literature on capture-recapture, where the issue of one-inflation has been recognised.

## 2. Methodology

According to the standard formulation, consider a closed population (no birth, death or migration) of size  $N$ . For each unit in the population, let  $Y$  be a random variable taking value  $j=0, 1, 2, \dots$  if the individual is observed/captured  $j$  times. We only observe the  $n$  individuals, such that  $j>0$ , with  $n \leq N$ .

Let  $p_j = \text{Prob}(Y = j)$  denote the probability of a unit being captured  $j$  times, and  $n_j$  denote the number of individuals observed  $j$  times, that is  $n_j$  is the frequency of the count  $j$ .

Our interest is to estimate the unknown  $n_0 = N - n$  units remain unobserved, actually the units observed in zero counts  $n_0$ , and consequently  $N$ , on the basis of some model for the observed  $n_j$ .

The Horvitz-Thompson estimator arises from the sum of both the unobserved and the observed cases  $n$ , by the solution of the estimating equation

$$N = Np_0 + N(1 - p_0) = n/(1 - p_0)$$

where  $N(1 - p_0)$  is the expected number of cases identified by the capture mechanism, estimated by the observed cases  $n$ . In the Horvitz–Thompson estimator, the probability of

being observed 0 times,  $p_0$ , needs to be estimated by an appropriate model that describe the observed zero-truncated counts.

Note that Bayesian inference for the population size  $N$  is straightforward under standard models for  $Y$ . Suppose for example that  $Y$  is Poisson distributed with mean  $\lambda$ . Let  $y = (y_1, \dots, y_n)$  be the observed number of captures for the  $n$  observed units and let  $s$  be their sum  $s = \sum_{i=1}^n y_i$

The likelihood function for  $\theta = (\lambda, N)$  is given by

$$f(y, n | \lambda, N) = \binom{N}{n} \exp\{-\lambda (N - n)\} \lambda^n \exp\{-\lambda s\}$$

Assuming a priori  $p(N) \propto 1/N$  and  $\lambda \sim \text{Gamma}(a_l, b_l)$  and setting  $n_0 = N - n$  the posterior for  $(\lambda, n_0)$  is given by

$$p(n, \lambda | y, n) \propto \binom{n_0 + n}{n} \exp\{-\lambda n_0\} \lambda^{n+a_l} \exp\{-\lambda (s + b_l)\} \frac{1}{n_0 + n}$$

which can be easily simulated via a Gibbs sampler. In fact, the conditional distribution of  $\lambda | n_0, y, n$  is a Gamma with parameters  $n + a_l, s + b_l$  and the conditional distribution  $n_0 | \lambda, y, n$  is a Negative Binomial with size parameters equal to  $n$  and probability  $\exp\{-\lambda\}$ .

To include the one-inflation in our model, we assume that in our population a specific behavioural mechanism is acting. That is, an individual that without that mechanism would face multiple captures, now has a positive probability  $\omega$  of being captured just once.

Let  $Y$  denote the observed number of captures for a unit, and  $Y^*$  the latent value we would observe without the behavioural mechanism. The two variables are linked by means of the following infinite transition matrix

$$P = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 0 & 0 & 0 & 0 & \dots \\ \hline 0 & 1 & 0 & 0 & 0 & \dots \\ \hline 0 & \omega & 1 - \omega & 0 & 0 & \dots \\ \hline 0 & \omega & 0 & 1 - \omega & 0 & \dots \\ \hline 0 & \omega & 0 & 0 & 1 - \omega & \dots \\ \hline \dots & \dots & \dots & \dots & \dots & \dots \\ \hline \end{array}$$

where the  $(k, j)$  -  $th$  element represents the conditional probability  $P(Y = j | Y^* = k)$ . Note that the first row comprises the probabilities  $\{p_{0j}\}_{j=0,1,\dots}$  and the first column the probabilities  $\{p_{k0}\}_{k=0,1,\dots}$ .

Let  $f(k|\theta) = P(Y^* = k|\theta)$  be the probability distribution, depending on some parameter  $\theta$ , of the number of captures without the behaviour effect, and let  $F(\theta)$  denote the associated c.d.f. Then, the resulting distribution for  $Y$  is the one-inflated model defined as follows:

$$P(Y = j) = \begin{cases} f(0|\theta) & \text{if } j = 0 \\ (1 - \omega)f(1|\theta) + \omega(1 - f(0|\theta)) & \text{if } j = 1 \\ (1 - \omega)f(j|\theta) & \text{if } j > 1 \end{cases}$$

The conditional distribution of  $Y^*$  when  $Y=j$  is concentrated on  $j$  when  $j \neq 1$ , while, when  $j=1$ , we have:



$$P(Y^* = k | Y = 1) = \begin{cases} 0 & \text{if } k = 0 \\ \frac{f(1|\theta)}{f(1|\theta) + \omega(1 - F(1|\theta))} & \text{if } k = 1 \\ \frac{\omega f(k|\theta)}{f(1|\theta) + \omega(1 - F(1|\theta))} & \text{if } k > 1 \end{cases}$$

This setting allows for different specifications for our count data  $Y^*$ , i.e.  $f(\theta)$  can be a Poisson density, as in Godwin and Böhning (2017), or alternatively  $Y^*$  can follow a Negative Binomial distribution, as in Godwin (2107). In both cases, a Gibbs sampler can be outlined for simulating from the conditional distributions of  $\omega, Y^*$ , and  $\theta$ .

### 3. Result

In this section, we apply the proposed Bayesian model to a selection of popular case-studies in capture recapture literature.

The following real cases are considered:

1. the number of prostitutes in Vancouver, presented in Rossmo and Routledge (1990);
2. opiate users in Rotterdam, already analysed by Cruyff and van der Heijden (2008);
3. heroin users in Bangkok, from Böhning et al. (2004) Viwatwongkasem et al (2008).

The evidence of one inflation in these data sets have been largely discussed already in Godwin and Böhning (2017) and Godwin (2017), where results from maximum likelihood analysis under Poisson and Negative Binomial distribution, respectively, can be found and compared with the results presented in this section.

Table 1 shows the results from our Bayesian models for the selected case studies, reporting the posterior modes and credible intervals of the population sizes,  $N$ , as well as the posterior medians of the model parameters. In the table, we indicate with  $\lambda$  the parameter of the Poisson distribution. For the Negative Binomial distribution, we adopt the size and probability parameterization, indicated with  $r$  and  $p$ , respectively. The one-inflated models are indicated with the suffix 'OI'. In table 1, we show also the results from a Bayesian analysis assuming the Poisson distribution and ignoring the one-inflation, to show the over-estimation occurring when one-inflation is ignored.

Table 1. The posterior mode and credible intervals for the population size  $N$ , posterior median for  $\omega$  and model parameters, for some popular real cases analysed

Prostitutes in Vancouver		N	HPD(N)	$\omega$	$\lambda$	r	p
	Poisson	1237	1178 - 1301		1.253		
	OIP	1016	980 - 1057	0.439	2.037		
	OINB	1304	1100 - 2174	0.274		1.947	0.604
Opiate users in Rotterdam		N	HPD(N)	$\omega$	$\lambda$	r	p
	Poisson	2929	2832 - 3038		1.174		
	OIP	2500	2418 - 2587	0.336	1.663		
	OINB	3769	3140 - 5533	0.086		1.397	0.618
Heroin users in Bangkok		N	HPD(N)	$\omega$	$\lambda$	r	p
	Poisson	9453	9427 - 9477		4.134		
	OIP	9364	9349 - 9380	0.207	5.004		
	OINB	10865	10629 - 11111	0.055		1.612	0.300

#### 4. Discussion and Conclusion

In this paper we propose a Bayesian approach to count data models with one-inflation, for use in Horvitz–Thompson estimator of the population size. The one-inflation can introduce severe overestimation in traditional capture-recapture estimator, if not properly accounted for. Our Bayesian approach results in population size estimates similar to the maximum likelihood ones, under non informative prior distribution, with the advantage of producing the credible intervals of the estimates as a by-product of the estimation procedure. Another advantage of the Bayesian approach is to allow incorporating previous knowledge on the hidden size of the population in the analysis, in the very favourable case in which this information is available to the analyst. The role of prior distributions is topic for further investigation.

We apply our Bayesian proposals to real data for estimating the size of some illegal population, using also some popular dataset available from the literature on capture-recapture, where the issue of one-inflation has been recognised. The properties of our model are highlighted also by a simulation study, not presented in these proceedings for the sake of brevity.

A possible extension of this work aims at incorporating, in a natural way, a Bayesian test for the one-inflation assumption.

#### References

1. Böhning, D., Suppawattanabodee, B., and Kusolvisitkul, W. and Viwatwongkasem, C. (2004). Estimating the number of drug users in Bangkok 2001: A capture-recapture approach using repeated entries in one list. *European journal of epidemiology*, 19, 12, 1075-1083
2. Böhning, D. and Friedl, H. (2021). Population size estimation based upon zero-truncated, one-inflated and sparse count data. *Statistical Methods & Applications*, 1-21
3. Böhning, D. and Kaskasamkul, P. and van der Heijden, PGM. (2018). A modification of Chao's lower bound estimator in the case of one-inflation. *Metrika*, 82, 3, 361-384
4. Böhning, D. and van der Heijden, PGM. (2019). The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain. *Annals of Applied Statistics*, 13, 1198-1211
5. Cruyff, M. J. and van der Heijden, P. G. (2008). Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biometrical Journal*, 50, 1035-1050
6. Godwin, R.T. (2019). The one-inflated positive Poisson mixture model for use in population size estimation. *Biometrical Journal*, 61, 6, 1541-1556
7. Godwin, R.T. and Böhning, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66, 2, 425-448
8. Rossmo, D., and Routledge, R. (1990). Estimating the size of criminal populations. *Journal of quantitative criminology*, 6 (3), 293–314
9. Viwatwongkasem, C., Kuhnert, R. and Satitvipawee, P. (2008). A comparison of population size estimators under the truncated count model with and without allowance for contaminations, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50, 6, 1006-1021