Check for updates

P. C. Mahalanobis in the Context of Current Econometrics Research*

Arnab Bhattacharjee
Heriot-Watt University, Edinburgh, UK
Tapabrata Maiti
Michigan State University, East Lansing, USA

It is a huge honour and privilege for us, the co-editors, to bring to the readership of *Sankhyā Series B* this Econometrics Special Issue celebrating the 125th anniversary of P.C. Mahalanobis. In the late 1980s and early 1990s, when we were passing through the premises of the Indian Statistical Institute, it was already more than a decade and a half after Mahalanobis, or "The Professor" as he is affectionately called by everyone at "The Institute", had passed away. However, it was impossible to miss The Professor's vision about The Institute, whether in the design of research and teaching, or the reminiscences and influences upon many of our teachers, or indeed the social environment at The Institute. Hence, when we were requested by Professor Sanghamitra Bandyopadhyay (Director, Indian Statistical Institute) and Professor Dipak Dey (Editor-in-Chief of *Sankhyā*) to edit this special issue, it was a matter of great honour but also a somewhat daunting challenge. We wish to provide a few qualifications at the outset.

First, Mahalanobis was a pioneering researcher in statistics and allied disciplines, but he was much more – a polymath, planner, educationist and visionary, and indeed one of the architects of the new Indian nation after independence. Together with his fundamental contributions in statistics, he also contributed significantly to research, thinking and societal value in planning and economics, not least in econometrics; here, our focus here lies exclusively on his contributions in econometrics. Significantly, he was the

---

*Throughout, "current" refers to the current period, say the first two decades of the 21st century. By contrast, "contemporary" refers to the period of Mahalanobis, approximately the 1930s to 1960s.

first Indian elected member of The Econometric Society and its first fellow elected from India (1951), and a founder of The Indian Econometric Society.

Second, while the Econometric Society defines its objective as "*the advancement of economic theory in its relation to statistics and mathematics*," in the context of this special issue, we have taken a narrower view of econometrics as the study of economic data aligned with economic reasoning and advancing the discipline of economics. Thereby, we offer a place of primacy to measurement and statistical inference based on economic data. We believe this focus is in line with the philosophy of Mahalanobis and the objectives of *Sankhyā*.[1]

Third, in the choice of topics and in inviting potential authors, we draw upon Rao (1963a, 1973), Rudra et al. (1996) and Kumar (1997, 2004). Further, we tried to maintain a balance between theory and applications and between leading experts and early career scholars, as well as representation across different regional and sub-disciplinary views. We were pleased and humbled by the overwhelming response from authors. With equally enthusiastic efforts by the reviewers themselves, the submitted papers were reviewed in the usual way and using customary quality benchmarks. Given the high quality and volume of submissions, it was soon apparent that what started as a single special issue would take two special issues to do full justice. The six papers collected in this *Special Issue on Econometrics in honour of P. C. Mahalanobis* represent the first part; the second special issue will appear in 2020.

The papers included here make very distinct contributions and, as a collection, nicely represent the context of current econometrics research. In their own distinct ways, each paper also represents in our view a current interpretation of the work and vision of Mahalanobis in econometrics. In the following paragraphs, we attempt to briefly highlight these connections along specific themes of The Professor's academic work (own research, and roles as the founding Editor of *Sankhyā* and the founder and Director of

---

[1]See, for example. In the Editorial to the first issue of *Sankhyā*, Mahalanobis (1933) wrote: "*We shall try to keep before us this comprehensive idea of the scope of statistics. We are convinced that statistics represents a fundamental method of analysis of data in the mass which is applicable to any science of observation, and we feel that it is desirable to emphasize this essential unity in the methodology of statistics*". In his own work, "*Mahalanobis made use of some simple mathematical models for studying some aspects of the problems concerned with planning. The models enabled him to isolate important sectors of the economy and study their mutual interdependence. ... The models developed do not follow the tenets of accepted economic theory ... but were developed in an atmosphere of pressing policy needs* " (Mukherjee, 1963). In this context Mahalanobis (1955a) wrote: "*I do not think that the [mathematical economics] models have any permanent value of their own. I have used them as scaffolding to be dismantled as soon as their purpose has been served.* "

the Indian Statistical Institute), professional work (as planner and official statistician) and as a visionary. In doing so, we have drawn liberally from the research of P. C. Mahalanobis, his interactions with leading economists and econometricians of his time (as evident, for example, from Rao (1963a)), and writings on The Professor's work and influence by other econometricians.

1 Official and Regional Statistics

Mahalanobis was appointed the Chairman of the National Income Committee in 1949 and was closely connected with the development of operational statistics in India in the early post-independence period. His work and thinking in terms of official and regional statistics were visionary.[2] They followed from his previous experiences in Bengal and were supported by the National Sample Survey, established in 1950, through the collection of socioeconomic sample data covering the entire country. Primary objectives were to provide information needed by the government for administrative purposes, including planning, and for the computation of national income. For planning purposes, he developed "*[w]hat is described in India as the single sector model of Mahalanobis ... a forward-looking Harrod-Domar type of model*" (Mahalanobis, 1950, 1952), which has investment, per capita income and its growth as core elements (Mukherjee, 1963).

As may be expected, Mahalanobis placed great emphasis on accuracy, on computation, and on timely compilation of national statistics; see, for example, Mahalanobis (1936a, 1963b) and Kuznets and Mahalanobis (1964).[3]

---

[2] "*[T]hinking about national income, in a way, is thinking about the performance of the nation as a whole. While working as the Chairman of the Committee, Mahalanobis became acutely aware of the national problems, national resources and allied matters. ... Also, even before this Mahalanobis undertook some fundamental work in the fields of regional and partial planning*" (Mukerjee, 1963).

[3] To quote: "*Indian official statistics have always been marked by a series of compromises, not only between "what is ideally desirable and what is actually obtainable", but also between statistical needs and administrative purposes. ... the collection and presentation of official statistics, which have therefore lacked consistency and completeness. These and the unavoidable but none the less regrettable delay in the publication of official statistics have been commented upon by a series of Committees and Commissions ... Practically all are unanimous in complaining about the delay in the issue of official publications. It is pointed out that the figures become completely out of date by the time they are published and thus do not serve any useful purpose*" (Mahalanobis, 1936a). "*I am recalling the two points made by him [R. A. Fisher], namely the need of cross-examining the data and the importance of computational work in statistics ... The only way to improve the quality of official statistics in India is by testing their accuracy in accordance with accepted scientific principles, ... [and] of ascertaining the margin of uncertainty in an objective manner*" (Mahalanobis, 1963b).

Together, a strong focus on reliable statistics at the regional level was apparent from the very beginning (Mahalanobis, 1933). Allan, Koop, McIntyre and Smith (2019, in this issue) makes an excellent current contribution to statistical accounts highlighting all of the above issues – focus on growth rate of regional income, computation, accuracy, timeliness and quantification of uncertainty, but based on statistical models – a great tribute indeed to the legacy of Mahalanobis. Specifically, Allan et al. (2019) use current methodology to nowcast economic growth in Scotland, which is classified in official statistics as a region of the United Kingdom, including nowcasts in "*pseudo real-time*," an important aspect of their model is also the use of mixed frequency data.

2   Mahalanobis Distance

Mahalanobis distance (or Mahalanobis D^2 statistic), originally developed in Mahalanobis (1927, 1936b), has been a popular and very useful measure of "closeness" of multivariate observations. This was a very fundamental contribution and much has been written about it from a statistical perspective; see, for example, Rao (1963b, 1973) and Rudra et al. (1996). Here we discuss some interesting connections with our current understanding of econometrics, focusing on contributions in this special issue. Mahalanobis distance has close links with entropy (and Kullback-Leibler as a generalisation) and other divergence measures. In turn, the maxent or minimum divergence (relative entropy) principle, also discussed in Mahalanobis (1950), has been very useful in econometrics for constructing measures of dependence, hypothesis testing of a parametric null against an omnibus alternative, estimation of conditional moments and specification testing; see for example, Parzen (1982), Robinson (1991) and Ullah (1996). In fact, Mahalanobis distance and similar concepts have been very useful as tools for modelling dependence and non-stationarity in time series and spatial data (Robinson, 2014). Hence, Mahalanobis distance offers a nice connection to several contributions in this special issue, namely: Bailey, Kapetanios and Pesaran (2019, in this issue); Balakrishna, Koul, Sakhanenko and Ossiander (2019, in this issue); Cai, Maiti, Bhattacharjee and Calantone (2019, in this issue); and Lee, Ullah and Wang (2019, in this issue).

Bailey et al. (2019) propose a new measure of the strength of cross-sectional (or spatial) dependence in panel data, together with asymptotic and finite sample performance and an application in finance. The proposed

measure is based on pair-wise cross-section correlations and is therefore related to the Mahalanobis D^2 statistic; the use of averaging, bootstrap and spatial thinking in this paper also bears the legacy of Mahalanobis, which we discuss later. Balakrishna et al. (2019) develop omnibus tests of a parametric linear autoregressive time series model with multiplicative errors. As discussed above, a common use of Mahalanobis distance and related divergence measures is in testing a parametric null hypothesis against an omnibus alternative, and this provides a nice contrast with the current approach developed in this paper. Cai et al. (2019) develop a Lasso-based model selection methodology for dependent data. Here, spatial dependence is modelled using geographic or geodesic distances, which offers an interesting connection with Mahalanobis distance; there is also a connection with nonparametric regression which we discuss later. Lee et al. (2019) makes a theoretical contribution to higher order asymptotics of an asymmetric least squares estimator, providing improved bias correction and mean squared error at very low and high percentiles. This has nice applications to risk measurement, but also a connection to Mahalanobis D^2 in its common use for identifying outliers. Such use is often compromised by inaccurate local measures of the covariance matrix, and in this context, the findings of this paper applied to the quantile regression model can be very useful.

3   Causal Models and Endogeneity

It is fairly apparent from his writings that Mahalanobis was primarily concerned with correlations rather than causal "*interactions*" (Mukherjee, 1963).[4] Nevertheless, his emphasis on forward looking planning models reflects a strong focus on counterfactual outcomes in the future that result

---

[4]For example, in Karl Pearson's obituary (Mahalanobis, 1936c), he quotes Pearson: "*I interpreted Galton to mean that there was a category broader than causation, namely correlation, of which causation was only the limit, and that this new conception of correlation brought psychology, anthropology, medicine, and sociology in large parts into the field of mathematical treatment. ... To him all science was 'description and not explanation'.*" "*[W]hat Mahalanobis has done is to give an indication of theory, and not any worked-out theory of economic development. ... Mahalanobis has not directly brought in the question of cause and effect into the picture. He has stressed more on simultaneity. ... Mahalanobis specifically admits of the interacting nature of the industrial and technological challenges, without pointing out which is prior*" (Mukherjee, 1963). Wold (1963) provided an excellent and illuminating discussion of The Professor's engagement with philosophical and analytical treatment of causation and its connections with least squares regression; see also Wold (1954, 1960), Simon (1955) and Lange (1963).

from today's planning decisions.[5] Such reasoning is quite closely in line with how causal modelling has subsequently evolved within the field of econometrics (and applied economics). Perhaps one cannot ignore the irony that in his own applied work Mahalanobis consistently eschewed causation in favour of correlation, and yet many of his contributions are found in the foundations of causal inference.

Baltagi and Ghosh (2019, in this issue) provides an excellent example of the theory and practice of modelling endogenous causal effects currently in econometrics. They consider treatment effects in a policy setting where the causal effect of a continuous treatment variable is measured by its impact on the marginal distribution of an outcome (*partial distributional policy effects*); however, the treatment itself is endogenous, which then requires new inference procedures. Some notable links are also there in the other contributions in this special issue. The system approach inherent in The Professor's planning models is integrated in the spatial Durbin model considered in Cai et al. (2019), from which some structural (causal) interpretations can be gleaned. Likewise, the (strong and weak) factor structure modelled in Bailey et al. (2019) can be provided structural interpretation, in line with the current practice of econometrics.[6]

The Professor's thinking on causation and economic theory was set in specific planning policy contexts. In fact, a central critique of contemporary economic theories considered by Mahalanobis was the fundamental focus, in classical economic theory, on aggregate demand, output, and consumption and not equally on distributional effects; see Stone (1963) for an excellent discussion including the context of Indian planning. Mahalanobis was intimately conscious of extreme inequalities in Indian society, and more generally issues of socio-economic equity and justice. This concern was also expressed in his work as the Chairman of the Committee on Distribution of Income and Levels of Living, and his development of Fractile Graphical Analysis (which we discuss later) can be partly related to his empirical studies on

---

[5]*"While Harrod-Domar models seek to describe how an economy moved in the past, and thus explain a phenomenon, Mahalanobis' models are essentially forward-looking planning models."* This is also apparent in The Professor's use of *"a simple simultaneous equation system and obtain his solution"* to his four-sector planning model (Mahalanobis, 1955a,b). Further, Mahalanobis (1961) emphasized that *"higher priority should be given to tiers or levels that are more slowly maturing,"* implying thereby that these are higher up in a notional causal ordering of the levels (Mukherjee, 1963).

[6]In his open letter to Mahalanobis, Wold (1963) also provides an excellent example of early conceptualisation of within and between effects in panel data regression. This too develops a nice contrast with its current econometric treatment, for example, in Bailey et al. (2019) in this special issue.

inequality (Mukherjee, 1973); see, for example, Mahalanobis (1959, 1963a).[7]
Likewise, Baltagi and Ghosh (2019) apply their proposed methods to a very
important and current problem of social justice, that of systematic patterns
of incarceration (imprisonment) in the United States.

4 Nonparametrics: Subsampling and Regression

In the later part of his career, The Professor made some fundamental
contributions to the early development of nonparametric regression. With
large data, the regression function of y on x can be calculated by averaging
all observations on y at each point x. With finite data, Mahalanobis (1958,
1960, 1963a) developed Fractile Graphical Analysis (FGA) which proceeds
by sorting x and partitioning into fractile groups, such as deciles, and then
computing and plotting conditional averages of y; essentially, this is an early
form of nonparametric regression based on a histogram sieve (Deaton, 1995).

In a way, focus on nonparametrics rather than parametric regression
models perhaps reflects The Professor's lack of comfort with structural and
parametric restrictions arising from economic theory. This is interesting be-
cause, notwithstanding errors inherent in parameter estimates and economic
measurements, Mahalanobis was more comfortable with using planning and
national accounts identities. Similar use of averages are also abundant in The
Professor's early work, for example in moving-block sampling methods for
spatial data, or interpenetrating subsamples; see, for example, Mahalanobis
(1946, 1950). In particular, Mahalanobis (1946) demonstrates careful spatial
analysis and a deep appreciation of J. A. Hubbard's early spatial sampling
methodology, both of which he built into his own method of block-sampling.
Further, in operationalising the single sector and two-sector planning mod-
els, Mahalanobis estimated key structural parameters using averages of time
series data across many countries "*to concretise his ideas on the rate of in-
vestment and rate of growth of national income in India*" (Mukherjee, 1963).
In fact, one might argue that his work laid the foundations for popular and

---

[7]To quote: in underdeveloped countries "*[a] very small group of families or persons have
the largest share of wealth, income and political and economic influence. In fact, the greater
the lack of economic development the fewer would be the number of persons who have the
effective power of making political and economic decisions. This makes it possible for a
foreign power to exert pressure on a small group of powerful persons to give concessions in
favour of the foreign power. Such arrangements, because they depend on the will of only
a small group of persons, are necessarily subject to violent changes from time to time* "
(Mahalanobis, 1959). He then argues that economic development is a necessary condition
for world peace. In fact, issues of social justice have concerned Mahalanobis throughout
his career, motivating him in line with the so-called "*Nehru-Mahalanobis ideology*" to view
"*statistics as a key technology for bringing about social change*" (Kumar, 1997).

useful resampling methods such as bootstrapping and subsampling that are now standard tools in econometrics and elsewhere (Hall, 2003; Robinson, 2014).

In this special issue there are several examples of current econometric treatments of spatial heterogeneity and spatial dependence (Bailey et al., 2019; Cai et al., 2019), which provides interesting contrast with The Professor's work. Likewise, the use of correlations (functions of second order moments, or sample averages) is central in the work of Bailey et al. (2019), which provides yet another link with The Professor's legacy. As discussed above, use of bootstrap is pervasive in econometrics today, and several different examples can be found in this special issue; see Bailey et al. (2019), Baltagi and Ghosh (2019), and Lee et al. (2019).

As a final point of comparison, regularization by assumption of sparsity is popular in the current theory and practice of econometrics. In this special issue, it is central to the measurement of cross-section dependence in Bailey et al. (2019) and likewise to high dimensional variable selection in Cai et al. (2019). This can be placed in contrast with traditional approaches prevalent at The Professor's time. There are two notable examples in Rao (1963a): macroeconomic models of national accounts identified by structural constraints (Frisch, with Parikh, 1963); and models of linear demand systems with explicit parameter restrictions (Stone, 1963). An interesting contrast also obtains from The Professor's own work on demand systems (Mahalanobis, 1963a; Bhattacharya and Mahalanobis, 1967), specifically on the consumption of cereals in India and regional disparities in household consumption in India, using his nonparametric FGA method and its generalisation.

Finally, we are grateful to all the authors, and likewise to all the reviewers, for their significant efforts and contributions, the Editor-in-Chief and co-editors of *Sankhyā* for continuous editorial assistance, and to the Indian Statistical Institute and its Director for institutional support. By its very nature, this preface is selective of certain aspects of Mahalanobis' work that, in our view, best capture the link with current econometrics research as represented in this special issue. Like any other selection, this is therefore inherently subjective. Nevertheless, we hope it provides a meaningful connection to the readership of *Sankhyā*, and places The Professor's influence upon econometrics in a current light. We believe that the six papers in this special issue make a significant contribution to the current literature in econometrics and are pleased to present this collection to the readership of *Sankhyā Series B*.

References

ALLAN, G., KOOP, G., MCINTYRE, S. and SMITH, P. (2019). Nowcasting using mixed frequency methods: an application to the Scottish economy. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-018-0181-2.

BAILEY, N., KAPETANIOS, G. and PESARAN, M.H. (2019). Exponent of cross-sectional dependence for residuals. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-019-00196-9.

BALAKRISHNA, N., KOUL, H.L., SAKHANENKO, L. and OSSIANDER, M. (2019). Fitting a pth Order Parametric Generalized Linear Autoregressive Multiplicative Error Model. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-019-00195-w.

BALTAGI, B.H. and GHOSH, P.K. (2019). Partial Distributional Policy Effects Under Endogeneity. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-019-00190-1.

BHATTACHARYA, N. and MAHALANOBIS, B. (1967). Regional disparities in household consumption in India. *J. Am. Stat. Assoc.* **62**, 317, 143–161.

CAI, L., MAITI, T., BHATTACHARJEE, A. and CALANTONE, R. (2019). Variable selection with spatially autoregressive errors: a generalized moments LASSO estimator. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-018-0176-z.

DEATON, A. (1995). Data and econometric tools for development analysis. *Handb. Dev. Econ.* **3**, 1785–1882.

FRISCH, R. and PARIKH, A.K. (1963). Parametric solution and programming of the Hicksian model. In: Rao, C.R. (1963a) (Ed.), *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday), Pergamon Press: Calcutta, India, 45–82.

HALL, P. (2003). A short prehistory of the bootstrap. *Statistical Science* **18**, 2, 158–167. Silver Anniversary of the Bootstrap.

KUMAR, T.K. (1997). An unfinished biography: Prasanta Chandra Mahalanobis. *Economic and Political Weekly* 7–13 June 1997, 1321–1332.

KUMAR, T.K. (2004). Why econometrics? *Economic and Political Weekly* 7 February 2004, 543–546.

KUZNETS, S. and MAHALANOBIS, P.C. (1964). Observations on national income estimates and related measures in India, January, 1957. *Sankhyā Series B* **26**, 3/4, 337–356.

LANGE, O. (1963). Elements of a Theory of System Behaviour. In: C. R. Rao (1963a) (Ed.), *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday), Pergamon Press: Calcutta, India, 137–152.

LEE, T.-H., ULLAH, A. and WANG, H. (2019). The Second-Order Asymptotic Properties of Asymmetric Least Squares Estimation. *Sankhyā Series B* (in this issue). https://doi.org/10.1007/s13571-019-00189-8.

Rao, C.R. (1963a). *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70 th Birthday), Pergamon Press: Calcutta, India. Available online at: https://archive.org/details/cr.rao.essays.on.econometrics.and.planning.

Rao, C.R. (1973). Prasantha Chandra Mahalanobis, 1893-1972. *Biographical Memoirs of Fellows of the Royal Society* **19**, 454–492.

Rao, C.R., Lahiri, D.B., Nair, K.R., Pant, P. and Shrikhande, S.S. (1963b). Scientific Contributions of Professor P. C. Mahalanobis. In: C. R. Rao (1963a) (Ed.), *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday), Pergamon Press: Calcutta, India, 321–342.

Robinson, P.M. (1991). Consistent nonparametric entropy based testing. *Review of Economic Studies* **58**, 437–453.

Robinson, P.M. (2014). Dependence and Nonstationarity in Time Series and Spatial Data. *Mahalanobis Lecture.* 8th Statistics Day Conference, Reserve Bank of India, Mumbai September 2014.

Rudra, A., Rao, B.L.S.P., Ghosh, J.K. and Bhattacharya, N. (1996). *Prasanta Chandra Mahalanobis: a biography*. Oxford University Press, Oxford.

Simon, H.A. (1955). Causality and econometrics: comment. *Econometrica* **23**, 2, 193–195.

Stone, R. (1963). Models for Demand Projection. In: C. R. Rao (1963a) (Ed.), *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday), Pergamon Press: Calcutta, India, 271–290.

Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference* **49**, 1, 137–162.

Wold, H.O. (1954). Causality and econometrics. *Econometrica* **22**, 2, 162–177.

Wold, H.O. (1960). A generalization of causal chain models (part III of a triptych on causal chain systems). *Econometrica* **28**, 2, 443–463.

Wold, H.O. (1963). A Letter Report to Professor P. C. Mahalanobis. In: C. R. Rao (1963a) (Ed.), *Essays in Econometrics and Planning* (Presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday), Pergamon Press: Calcutta, India, 309–320.

Arnab Bhattacharjee
Spatial Economics & Econometrics
Centre (SEEC),
Heriot-Watt University
Edinburgh, UK
E-mail: a.bhattacharjee@hw.ac.uk

Tapabrata Maiti
Department of Statistics and
Probability,
Michigan State University
East Lansing MI, USA
E-mail: maiti@msu.edu