iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: IV Month of publication: April 2020

DOI: http://doi.org/10.22214/ijraset.2020.28079

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

Robust Frontier Estimation from Missing Data

Khalid Ul Islam Rather[1], Rizwan Yousuf[2]
[1, 2]*Division of Statistics and Computer Science, Main Campus SKUAST-J, Chatha Jammu-180009, India.*

Abstract: The main objective of this study is to identify and deal missing information by using the method of robust frontier. The existence of missing observations in the estimation problems present in random sampling can be considered inconsequential. Although the hazard of misunderstanding is high because the non-responses may be generated by the existence of a very different behaviour of a group of units. This is particularly important at what time human populations are sampled. Also The problem of imputing missing observations under the linear regression model is considered. Throughout the study regression model is used to impute the values in the place of missing information. The robust frontier estimation method used to deal with missing information or non-response. The purpose to estimate robust frontier to fill up the non-response observation by using regression imputation. Efficiency comparisons is made by observing the different properties.
Keywords: Missing data, Non-response, Regression Model, Robust Frontier Estimation Method and GLM Efficeincy.

I. INTRODUCTION

The most common problem in surveys is non-response or missing information. Missing information is common problem now days, which cannot be reduced even by increasing the sample size. Imputation is used in survey sampling to handle non-response or missing by assigning values to missing data set. A study conducted by Hansan and Hurwitz (1946) assumed that a sub sample of initial non respondents is re-contacted with a more expensive method; they suggested the first method by mail questionnaire and the second method by personal interview. Non-response occurs when sampling unit select for a sample in survey are not interviewed because they are unavailable, un-willing to participate in the survey due to some reasons. The relationship between unit and item non-response is studied by (Dixon et al., 2002).

Non-response always exists when surveyed population did not respond, when an eligible sample units fail to response in a survey due to some reasons is known as unit non-response. When a respondent unit does not provide useful answer to particular item in a questionnaire is known item non-response, see (De Luca and Peracchi, 2007).Unit non-response and item non-response create many problems in analysis and for the analyst it can increase the sample variance. There are many methods to handle the problem of item non-response and unit-response (Rauda and Gunzaiez, 2008). Ismail et al., (2015) discussed that non-response reduces the efficiency of survey.

II. IMPUTATION PROCESS:

This is a process of replacing missing data with some values. There are some rules for replacing the values in place of missing observation. Imputation is used in survey sampling to handle non- response by assigning values to missing data set and producing complete data set for further use and analysis, see (Brick et al., 2005). Imputation is most widely used in sampling surveys when data is missing. There are different methods of imputation available. These incorporate weighting procedures, single attribution, and various ascriptions (MI). MI plans to create conceivable attributions for the missing qualities, to precisely reflect vulnerability, also, to protect critical information connections and parts of the information appropriations. The greater part of the examinations have ascribed missing information by taking into account the example of missing (missing totally indiscriminately or missing at irregular; monotone or non-monotone missing information), kind of credited variable (continuous or discrete) and techniques.

III. SIGNIFICANCE EﬃCIENCY

Current study information investigations have been effectively utilizing MI method to address missing information issues too drawing deductions on parameter coefficients. In this paper, the significance of representing inefficiency on MI of missing information is assessed by looking at the stochastic wilderness examination (SFA) too summed up straight models (GLM) factual methods. Second, the significance of MI on specialized efficiency measures is assessed utilizing SFA. Specifically, the significance of MI on SFA specialize efficiency assessed under option distributional assumptions-half typical truncation and and exponential is assessed.(Daouia,2018).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

IV. ROBUST FRONTIER ESTIMATION METHOD

The problem of missing information or non-response is carried throughout the study. If missing information in the observations is ignored then sample size becomes smaller. Then by increasing sample size the efficiency of estimators can be improved. By Robust Frontier method of estimation we remain try to fill up the missing observation and non-response. When fill up the non-response observations then with same sample size estimators give more precise results. The regression model is as follow:

$$Y=X\beta+Z\gamma+ \varepsilon$$

The piece and measurement of Z relies upon the utilization of time arrangement (TS), board irregular effects (PRE) or various levelled direct model (HLM) measurable technique. For instance, consider a three-way board arbitrary effects show that incorporates three different factors. The three elements are salary gatherings, area and nation, furthermore, are treated as free arbitrary factors. Be that as it may, the three-way board demonstrates does not consider the normal qualities that every nation shares inside a specific locale. Interestingly, three-way various levelled straight model does consider the shared traits that emerge as a result of the settling or various levelled structure of nation inside a locale and salary gatherings. Thus, two-way board arbitrary effects differ from two-way progressive direct model with deference to various levelled structure of the two variables - nation and locale. In the event that Z network is set to zero, condition 1 comes down to a TS factual strategy.

V. RESULTS

In 1977, Aigner Lovell and Schmidt, Meeusen and van sanctum Broeck, and Battese furthermore, Corra all the while presented the stochastic wilderness demonstrate that represents the mistake term, ε into a symmetrical irregular blunder, v and an uneven mistake or on the other hand inefficiency, u. The ordinary half typical and an exponential dissemination was as- sumed by Aigner, Lovell and Schmidt (1977), while Meeusen and van cave Broeck (1977) accepted an exponential appropriation of the inefficiency term In 1982, Jondrow, Materov, Lovell and Schmidt recommended a strategy to es- timate firm specific inefficiency measures. The stochastic outskirts model can be used to speak to a Cobb- Douglas creation work as where yt speaks to a 1×T grid; xt speaks to a K×T lattice of exogenous info amount and time incline factors with T speaking to the worldly (time arrangement) measurement; α is the catch, β is the related parameters of information amount factors; and ε speaks to a 1×T network of unadulterated arbitrary blunder and disintegrated into v speaks to the arbitrary blunder and v ~ N (0,σ2 v), u speaks to the contrarily skewed uneven inefficiency and can be spoken to with elective circulations counting half ordinary, exponential, or truncated typical dispersion. Subtle elements for the dissemination can be found in the arrangement of 1977 articles. We signify the recurrence of these cases, separately, by n(xik |?), n(?|πij), and n(?|?). Consider, for example, the Bayesian networking and suppose that we wish to estimate the contingent likelihood of X3 = 0 (named x31 in figure 1), given the parent factors configurations 3 = (1, 0) (which we called π33 in figure 1), from the fragmented informational collection in Table 1. The cases a1 and a6 are finished and decide n(K31 |π33) = 2. All cases a3, a5, a7, a8, and a10 can be reliably finished as K3 = 0|3 = (1, 0). The case c3 decides n(?|π33) = 1. The cases a7 and a10 decides n(K31 |?) = 2, while the two cases a5 and a8 decide n(?|?) = 2. By finishing the cases a3, a5, a7, a8, and a10 as K3 = 0|3 = (1, 0), we make a specific predictable culmination of the informational collection, in which the occasion K3 = 0|3 = (1, 0) happens the biggest number of times. This thought is the instinct behind the definition of the virtual recurrence n(Kik |πij). The amount n(Kik |πij) is the most extreme number of deficient cases (Ki,I) that can be reliably finished as (Kik,π ij) and is defined by

n(Kik |πij) = n(?|πij) + n(Kik |?) + n(?|?).

Case	K_1	K_2	K
a_1	1	0	0
a_2	0	?	1
a_3	1	0	?
a_4	?	?	1
a_5	1	?	?
a_6	1	0	0
a_7	?	0	0
a_8	?	?	?
a_9	?	0	1
a_{10}	?	0	0

Presently take note of that, on the off chance that we finish the cases a3, a4, a5, a8, and a9 as K3 = 1|3 = (1, 0), we make a steady fruition of the informational index in which the occasion K3 = 0|3 = (1, 0) happens the base number of times. We then define the virtual recurrence nKik |πij as the most extreme number of inadequate cases (Ki,I) that can be attributed to πij without expanding the recurrence n(Kik |πij), or, in other words

n(Kik |πij) = n(?|πij) + h=kn(Kih |?) + n(?|?).

VI. CONCLUSIONS

In this study we fill up the missing information or non-response by using Robust Frontier method. Regression model is used to estimate the Robust Frontier method to fill up MI. Assume currently that we have a tendency to want to utilize a Bayesian system, quantified with assessments registered from a fragmented informational index, to foresee the estimation of the variable Xi, on condition that we have a tendency to watch the estimations of a set of alternate factors within the system. The arrangement of variable qualities watched is named proof,which we have a tendency to mean bye .The arrangement is to register the chance conveyance of Xi given the proof utilizing some commonplace unfold algorithm(Pearl, 1988; Castillo, Gutierrez, and Hadi, 1997) and subsequently to decide on the estimation of Xi with the largest chance, given e. we are able to conjointly proliferate the chance interims registered by the RBE with one amongst the present unfold calculations for interim primarily based Bayesian systems and figure a chance interim [p(xik |e) ¯ p(xik |e)] for every esteem p(xik |e). Such interims is utilized to form a forecast that doesn't depend upon a particular presumption regarding the model for the missing info. we have a tendency to come through this assignment by choosing a live where at base the selection of the Xi esteem. The random strength paradigm (Kyburg, 1983) chooses the esteem xik of Xi if the bottom chance p(xik |e) is greater than the best chance ¯ p(xih |e), for any h = k. random predominance is that the most secure and most moderate basis since the forecast is freed from the dissemination of missing info.

REFERENCES

[1] Dixon, J. A., Mahoney, B., & Cocks, R. (2002). Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt. Journal of Language and Social Psychology,21(2), 162-168.

[2] Hansen, M. H. and Hurwits, W. N. (1946). The problem of non-response in sample surveys,Journal of the American Statistical Association, 41, 517 – 529.

[3] Rueda, M., & González, S. (2008). A new ratio-type imputation with random disturbance. Applied Mathematics Letters, 21(9), 978-982

[4] De Luca, G., &Peracchi, F. (2007). A sample selection model for unit and item nonresponse in cross-sectional surveys.

[5] Brick, J. M., Jones, M. E., Kalton, G., & Valliant, R. (2005). Variance estimation with hot deck imputation: a simulation study of three methods. Survey Methodology, 31(2),151.

[6] Ismail, M., Hanif, M., &Shahbaz, M. Q. (2015). Pak. J. Statist. 2015 Vol. 31 (3), 295-306 generalized estimators for population mean in the presence of non- response for two-phase sampling. Pak. J. Statist, 31(3), 295 306.

[7] Skinner, C. J., & Rao, J. N. K. (2002). Jackknife variance estimation for multivariate statistics under hot-deck imputation from common donors. Journal of Statistical Planning and Inference,102(1), 149-167.

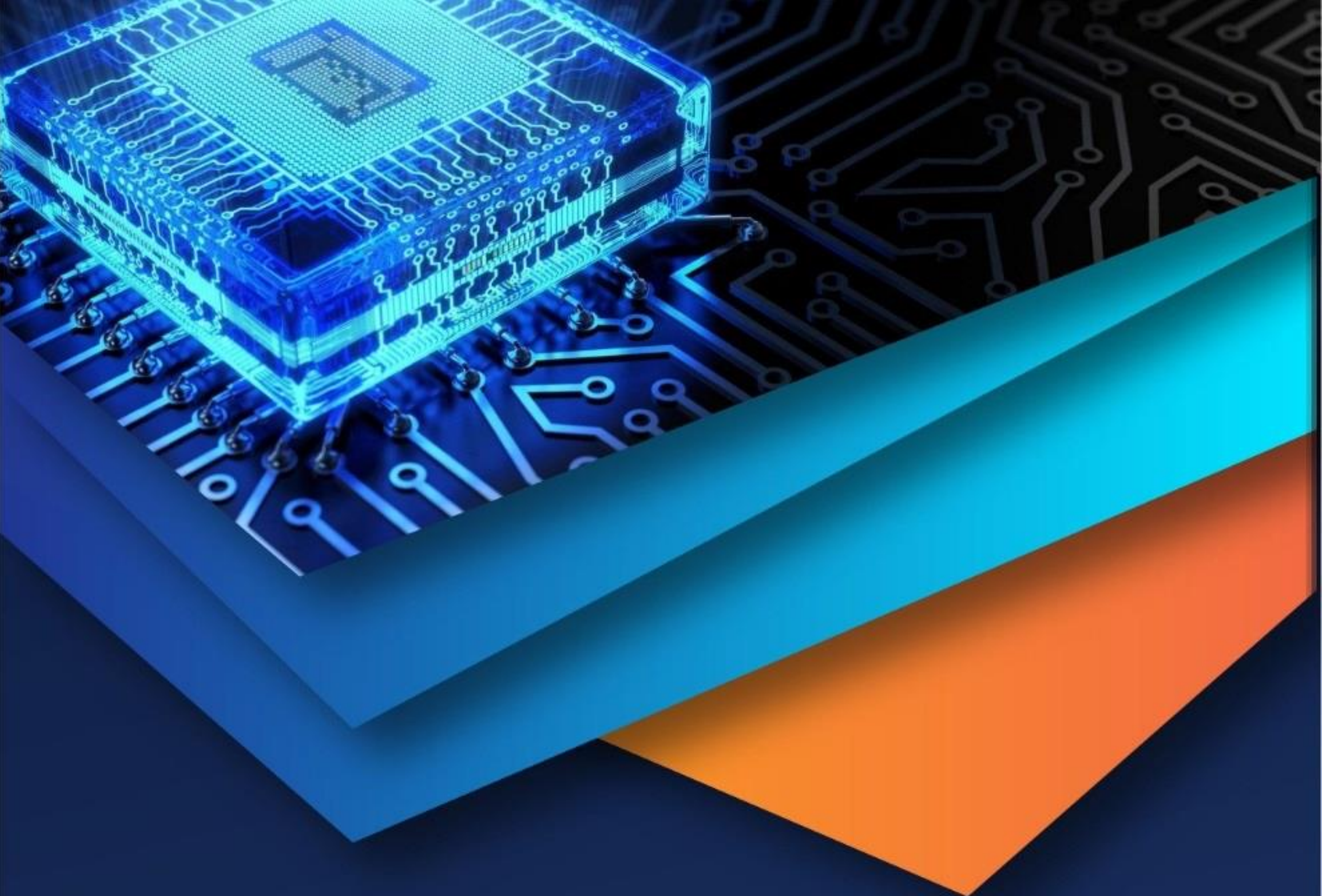
[8] Kalton, G., &Kasprzyk, D. (1982). Imputing for missing survey responses. In Proceedings of the section on survey research methods, American Statistical Association (Vol. 22, p. 31) American Statistical Association Cincinnati.

[9] Srivastava, M. S., &Dolatabadi, M. (2009). Multiple imputation and other resampling schemes for imputing missing observations. Journal of Multivariate Analysis, 100(9), 1919 1937.

[10] Singh, R., Verma, H. K., & Sharma, P. (2016). Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations. Journal of Modern Applied Statistical Methods, 15(1), 19.

[11] Daouia, A. (2018). Robust frontier estimation from noisy data: a Tikhonov regularization approach. Retrieved from ideas: https://ideas.repec.org/p/tse/wpaper/30543.html

[12] RAMONI, M. (2001). Robust Learning with Missing Data. Bayesian learning, Bayesian networks, 1-24.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY