

PS7 DScourse 2024

Emilien Akotenou

March 26, 2024

1 Summary Statistics

Table 1: Summary Statistics of Variables

	logwage	hgc	tenure	age
Minimum	0.0049	0.0	0.000	34.00
1st Quartile	1.3623	12.0	1.583	36.00
Median	1.6551	12.0	3.750	39.00
Mean	1.6252	13.1	5.971	39.15
3rd Quartile	1.9362	15.0	9.333	42.00
Maximum	2.2615	18.0	25.917	46.00
NA's	560			

1.0.1 Missing Data

The log wage variable is likely Missing at Random (MAR). The missingness seems to be related to other observed variables like education, age, and job tenure. It is unlikely to be MCAR since the probability of missingness appears correlated with these observed characteristics. However, there is no clear reason to think the missingness depends on the unobserved wage values themselves after controlling for the other variables, so MNAR is also unlikely.

2 Results

2.1

The true value of the years of schooling coefficient is 0.093. The listwise deletion model assuming MCAR underestimates this at 0.073. Mean imputation and single regression predictions assuming MAR come much closer at 0.089 and 0.091. Multiple imputation yields an estimate of 0.087, also quite close to the true value.

This demonstrates that assuming the data are MCAR and dropping incomplete cases can lead to biased coefficient estimates if that assumption is violated. If the data are actually MAR, methods that leverage the observed information like regression prediction or multiple imputation perform significantly better. With a large amount of missing data, even a MAR violation can meaningfully impact results.

3 Project Update

I plan to write about 'Infrastructure's Contribution to Socio-economic Development,' with a specific focus on Internet infrastructure. I intend to explore how internet infrastructure and digital technologies impact (i) health outcomes and (ii) labor reallocation and firm productivity, and how they lead to (iii) better public service provision in developing countries.

For this project, I am developing an idea related to Internet roll-out and child mortality in Kenya. In 2013, Kenya developed a national plan strategy aiming to provide access to the internet for all

Table 2: Regression Results for Different Imputation Methods

	MCAR	Mean Impute	MAR	MICE
Years of School	0.062 (0.005)	0.050 (0.004)	0.057 (0.004)	0.062 (0.005)
collegenot college grad	0.145 (0.034)	0.168 (0.026)	0.145 (0.026)	0.145 (0.034)
tenure	0.050 (0.005)	0.038 (0.004)	0.041 (0.004)	0.050 (0.005)
I(tenure^2)	-0.002 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.002 (0.000)
age	0.000 (0.003)	0.000 (0.002)	-0.001 (0.002)	0.000 (0.003)
marriedsingle	-0.022 (0.018)	-0.027 (0.014)	-0.025 (0.014)	-0.022 (0.018)
Num.Obs.	1669	2229	2229	
R2	0.208	0.147	0.192	
F	72.917	63.973	88.130	

health facilities and 35% of households by 2017. I am curious about whether internet infrastructure leads to a reduction in child mortality. If there is any effect, I will investigate its channels: The internet facilitates healthcare extension, improves communication, and enables mobile money payment. I will explore how these factors contribute to the reduction in child mortality.

Following my discussion with you on the data issue related to my topic, Internet roll-out and child mortality in Kenya, I started looking at countries with panel data on related topics such as internet usage, health, labor market, and other outcomes. I found an individual, household, and firm unbalanced panel dataset from Nigeria collected by the World Bank's Living Standards Measurement data.

Additionally, through my literature review, I came across two papers closely related to my research interests: 'The Welfare Effects of Mobile Broadband Internet: Evidence from Nigeria' by Bahia et al. (2020), published by the IZA Institute of Labor Economics, and 'The Arrival of Fast Internet and Employment in Africa' by Jonas Hjort and Jonas Poulsen (2019), published by the American Economic Association (AEA). I also found 'Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings' by Greenstone, M., Hornbeck, R., and Moretti, E. (2010), published in the Journal of Political Economy.

I have written a referee report on the first two papers and plan to replicate their results while also identifying potential areas where I can contribute to the existing literature.