1. Short description of the paper and its causal inference findings;

The paper that was selected for the purpose of this referee report aims to find any statistical justification for existence of wage gap between white and black ethnicities. The paper is a work of Neal and William (1996) that wanted to research any signs of black and white wage gap that will show discrimination of labor market. The novelty of their research lies in the fact that as Neal and William (1996) claim all previous researches focused on productivity as the main factor to explain for differences in wages among black and white people, but, the issue was two tailed where the first tail focuses on endogeneuty of the main explanatory variable that explains wage gap, while the second tail focused on the wrong choice of proxy variable that controls for work skill.

When reviewing the first issue in details, it is seen that majority of previous papers focused on the explanation of wage gap through work skills and productivity while these variables were already affected by labor market discrimination that suggested better education, work skills and additional education to white people. In other words, as people join labor market black people had less chances to receive and encounter education and training that was widely suggested for white people. Thus, in their paper, Neil and William (1996) decided to omit any such endogenous explanatory variables and use only exogenous variables to explain the wage gap.

The second issue was that majority of explanatory variables used in previous studies relied proxy variables that failed to measure productive skills. All proxies included factors such as education, parent education, training that have already been contaminated by discrimination. Thus, they have found a new variable (that was proven by other research as a good measure of productive skills) called AFQT (Armed Forces Qualification Test) that significantly measures human capital (work skills) that that was nurtured by people until the age of 16-18. They believe using a proxy for worker skill that is not contaminated by labor marked discrimination is the best approach.

Consequently, several models were used in order to see if there is a statistically significant wage gap between black and white. Using OLS regression with log wage as the dependent variable and independent variables such as race, age, schooling for two different genders (men, women) and they found that firstly, there is a statistically significant difference in skills among black and white and this gap have grown, and secondly there is a statistically significant wage gap between black and whites among males (-24.4%) and females (-18.5%). Interestingly, they have applied AFQT in order to adjust for worker skills differences and found that wage gap due to race drops to 7.2% for males and 3.5% for females. This, as authors claim, have showed that AFQT explained for 33% of variation in wages due to racial discrimination. However, the most interesting part is that further regressions have proved that return to pre-labor market education (until they reach 16-18 ages) is higher for white people when compared to blacks.

Finally, the paper ended up with a conclusion that race affects both wage gap and worker skills. Thus, the government should address the issue that black children acquire less worker skills.

2. Review on its research design and the application of causal inference method.

Methodology.

This paper relies on OLS regression in order to find causal relationship that explains differences between wage via race. Authors performed several regressions to control for various factors, but mainly the model included a dependent variable in the form of log(Wage) and later Wage and several independent variables such as race, schooling, age, AFQT. It worth mentioning that regressions were performed for males and females separately. There is no sufficient explanation for this, but it gives a feeling that this was done in order to avoid inconsistency in causal inference when it comes for females (Hispanics earning more than Black and White.

The methodology was novel and added to existing approaches in two ways. The first first improvement lies in the fact that a new proxy variable for skills was used. Instead of GPA, years of schooling, parent years of schooling, they have used AFQT test scores to control for worker skills gained before the entry to the labor market. The research argues that this variables is best and has no contamination with racial discrimination since it was already proven in the research performed by National Academy of Science and Department of Defense. The second improvement was that they considered worker skills gained after entering the labor market as endogenous variables and avoided this mistake.

Data.

This paper uses a panel dataset of nearly 12.5 thousands people with a date of brith between 1957 and 1964. Furthermore, they have reduced this sample to younger generation who have not joined the labor market before 1980 by explaining that this way they will avoid contamination

Variables.

As it was mentioned before, this paper claims to identify an endogeneuty issue in the form of proxy variable for worker skills and productivity. At the same time, they believe that a wrong choice of proxy variable for worker skills and productivity can cause biased results. Consequently, they suggested the use of AFQT as a proxy for worker skills and claim it is unbiased. This was they believe they can attain a tru causal relationship between race and wage gap.

3. Weakness of the paper from methodological perspective

There are many weak points in this paper that can potentially affect the validity of claims produced in this paper. The first main mistake, is that we can not believe all these inferences and statistical significance until all MLR 1-5 assumptions (Gauss-Markov assumptions) are validated (MLR 6 is not vital in this case due to huge sample size). Failure to validate MLR 1-5 assumptions can lead to biased and inconsistent estimators.

The second Achilles' Heel of this paper is that sample. The first issue with the sample is that it fails to comply random sample concept to some extend. By reviewing data collection we see that sample consists of only people with AFQT score, which is not a good representation of the whole population. Also, authors mention that the sample is overrepresented by black and hispanics in population, which again makes sample bad.

The final issue is that they consider two different regressions: for men and women separately. This methods opens a door for subjective approach to data and statistical inference.

Reference list:

Neal, Derek A., and William R. Johnson, 1996. "The Role of Premarket Factors in Black-white Wage Differences." *Journal of Political Economy* 104(5): 869-895.