1. **METHODOLOGY**

A concern in modeling the correlation between political participation and remittances is that individuals who receive remittances are not randomly selected. In that regard, the political effect of remittance could in part be driven by unobserved cofounders that could be systematically associated with selection into the remittance-receiving category and with subsequent political engagement (Ebeke & Yogo 2013). When endogeneity problems exist, Ordinary Least Squares (OLS) regression normally produces biased and inconsistent estimates (Tsiboe et al 2021; Ebeke & Yogo 2013). Matching has become a popular and effective statistical technique for reducing selection bias in the social sciences research (Abadie & Imbens, 2006; Dehejia & Wahba, 1999; Sekhon, 2004; Imai, 2005; Tsiboe, Egyir & Anaman, 2021). These techniques are useful in observational studies where treatment is non-random. In a nonrandomized study, it can be hard to compare treatment and control groups as they may both differ significantly based on observables. Matching provides a way to compare the outcomes of the treatment with those of matched non-treatment units based on similarities in covariates (Rosenbaum, 2005). This provides a treatment and matched non-treatment group that is sufficiently similar such that any confounder effects are eliminated. By achieving balance on the covariates, matching will prove useful in purging the inconsistency and biasness caused by the potential endogeneity problem by separating the effect of remittances from other factors shaping respondents’ electoral and non-electoral behaviors. An advantage of matching is that they do not require the investigator to make a heroic functional form for the outcome, as such, model dependence is greatly reduced (King & Nielsen 2019; Tsiboe et al. 2021).

Different matching techniques have been proposed in the causal inference literature. The more common one is the propensity score procedure proposed by Rosenbaum & Rubin (1983). They argue that correctly specified propensity score models will asymptotically remove the bias due to confounders. The procedure involves matching treatment units to control observations using the propensity score difference as a distance measure. A few papers in the remittances literature such as Esquivel & Pineda (2007), Cox-Edwards & Rodríguez-Oreggia (2009), and Ebeke & Yogo (2013) have utilized propensity score matching techniques. Despite the popularity of propensity score matching, the procedure is fraught with fundamental problems. King & Nielsen (2019) argue that propensity scores should not be used for matching. This is because the procedure attempts to approximate a completely randomized experiment rather than approximating a fully blocked random experiment which is more efficient. As a result, the propensity score matching technique is more likely to increase imbalance, model dependence, inefficiency, and bias in the estimation of treatment effects. Drake (1993) also argues that a mis-specified propensity score model produces bias estimates. Heckman et al. (1998) show that even when the right propensity scores are known, they still do not do a good job of reducing the variance of the estimator.

Traditionally, propensity score matching relies on manual optimization by iteratively matching observables and rechecking balance. This manual process often leads to suboptimal solutions (King et al., 2017). To avert this problem, we utilize genetic matching, which is unique amongst the new matching methods in balancing covariates.

* 1. **Why Genetic Matching in this Study?**

Genetic matching was first introduced by Hollard (1992) and advanced by Diamond & Sekhon (2013). This technique is a more efficient way of achieving covariate balance compared to the procedure suggested by Rosenbaum & Rubin (1984) which involves iterative estimation of propensity scores to achieve balance. The procedure is a generalization of propensity score and the Mahalanobis distance matching techniques that makes use of a search algorithm developed by Mebane and Sekhon to obtain covariate balance across matched control and treatment groups (Diamond & Sekhon, 2013). The idea is to match treated observations to control observations with the smallest weighted Mahalanobis distances (MD). The main problem with the use of standard MD methods is that they don’t perform very well when we have covariates with nonellipsoidal distributions. The genetic matching technique introduces innovations by the use of an evolutionary search algorithm that identifies the appropriate weights for each covariate to achieve covariate balance even under nonellipsoidal distributions (Sekhon & Grieve, 2012). This is done by identifying the most imbalanced covariates and attempting to improve their balance gradually in successive iterations (Henderson & Chatfield, 2011). The generalized Mahalanobis distance measure for two units *i* and *j* is stated by Diamond & Sekhon (2013) as:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where *X* is a matrix of observed confounders, *S* is the covariance matrix of *X* and *XT* (The transpose of *X* ). *W* is a positive definite matrix of weights and is the Cholesky decomposition of *S.*

It is reasonable to think that the covariates in this study should not be given equal weights. Their potential confounding effect on treatment assignment will be different [Are you 100% sure with what you are saying?]. The use of different weights by the genetic matching technique is a good way of attaining covariate balance for this study.

In the proceeding exposition, we lay out the matching framework in application to our data.

In this paper, remittance is our treatment variable. Let be a dummy variable equal to 1 if an individual received remittances [ You are not clear with the treatment and control group. You have to categorically state the treatment group and control group. Don’t assume readers know what you are doing. ] We define as the outcome of political participation measure *j* of an individual *i* who received remittances and as the outcome of political participation measure *j* of an individual *i* who did not receive remittances. We let be a vector of pretreatment covariates. For identification, we make a conditional ignorability or exchangeability assumption. That is we assume that the individual’s potential outcome as regards political participation is independent of his potential treatment assignment, conditional on the pretreatment covariates .That is:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

This implies that upon controlling for the pretreatment covariates, the selection into treatment is approximately random. By conditioning on , we ensure that any observed confounders do not affect treatment and control assignments. We attempt to control for most covariates that we believe could potentially affect the reception of remittances. This assumption will be valid if our sensitivity analysis shows that our treatment effects are not very sensitive to unobserved confounders.

We implement the genetic matching procedure to build a representative control group. We then proceed to estimate the treatment effects. The idea is to compare the outcomes of political participation of an observation in the treatment group to the political participation outcomes of its matched counterpart in the control group. We concentrate on the Average Treatment Effects on the Treated (ATT). Rather than average over the entire sample, we average over the treated group as this is more representative. The ATT is presented as:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

The outer expectation implies that the ATT is conditioned on the covariate profiles in the treatment category. The estimate of the ATT is computed as:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Where is the estimated treatment effect of remittances on a political participation measure *j.*

There are other innovative matching techniques in the literature that we make use of as a comparison with the genetic matching technique. Each of these has its advantages and potential problems. However as our results will show the genetic matching technique gives us better covariate balance relative to these methods, hence our decision to stick to this method.

We employ other matching techniques such as the nearest neighbor matching, optimal matching, and full matching. The nearest neighbor matching is the easiest to implement. This technique involves matching each treatment observation with the closest eligible observation in the control group based on a distance measure. The commonest distance measure used is the propensity score difference between treated and control units. It is also called greedy matching because each pairing is done without reference to how other units have been paired. This poses a potential problem as the quality of matching is affected by the order in which the treated observations are matched. Another common concern with this technique is that it could result in poor matching, especially when there are no control observations that have propensity scores similar to a treated observation. Methods such as caliper matching is often employed to prevent this problem. In this case, a match is selected only if it meets a certain caliper (Stuart, 2010).

Optimal matching improves on the nearest neighbor matching in that it takes into account all matches in choosing matches for each treatment observation. It makes use of a global distance measure. Gu & Rosenbaum (1993) show that full matching distances are smaller compared to greedy matching when the number of control variables are few. They also argue that even though this technique does better in reducing the distance between treatment and matched control groups, it doesn’t do a better job at creating groups with good balance. In terms of the selection of control observations, these two techniques are quite similar. Optimal matching however does a better job at assigning these controls to treatment units.

Full matching is an improvement on optimal and nearest-neighbor matching techniques. This technique creates matched sets. Each matched set contains at least one treated observation or at least one control observation. The creation of these sets means that all observations in the data are utilized. The matched sets are generated in an optimal way such that treated observations that have many similar comparison observations based on propensity scores are grouped with many comparison individuals and treated observations with less similar comparison observations are grouped with fewer comparison individuals. This technique ensures that all the data is utilized while achieving balance based on the propensity score.

As mentioned earlier we will see that the genetic matching technique attains better covariate balance compared to these techniques.