



UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

SCHOOL OF RESEARCH & POSTGRADUATE STUDIES Dissertation Correction Compliance Form (Post Viva Form)

Date: 14 September 2026

Name of Candidate: Bob Robert Tumushiime
No: B31332

Reg. No: J25M19/009 Student

Title of Dissertation: *Machine Learning Decomposition of Gender and Rural-Urban Wage Gaps: Evidence from Uganda National Household Surveys*

Note on the title: the title above is the revised title adopted in response to comment 3 below. The title at examination was "Explainable Machine Learning for Wage Gap Decomposition: Evidence from Uganda National Household Surveys (2012/13 - 2023/24)".

A. COMMENTS BY EXTERNAL EXAMINER

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Some allegations or claims in the problem statement are unsupported.	Section 1.3 was rewritten so that every claim now has a source. The female-to-male wage ratios of 0.73 urban and 0.67 rural come from Kagundu and Pavlova (2007), the overall gap of about 39 per cent to Sebaggala (2007), the monthly gender gap of about 32 per cent to UN Women (2024), and the Kampala-Karamoja consumption and poverty figures to UBOS (2025). Claims that no prior study exists now refer only to the literature reviewed in Chapter Two and point to Table 2.2, instead of being stated as absolute facts. The unsupported claims that the gaps are "poorly understood" and that OLS "cannot capture" nonlinear dynamics have been removed and replaced with a cited argument about model form, drawn from Lemieux (2006), Psacharopoulos and Patrinos (2018), and Kagundu and Pavlova's own public-private ratio of 1.04 against 0.55.	Section 1.3, p. 4

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
2	The abstract is inadequate and should clearly address what was done, why it was done, how it was done, what was found, and the significance of the findings.	The abstract was rewritten from scratch. It now covers all five points in the order asked for, one per paragraph: paragraph one gives the problem and why the work was needed, paragraph two what was done and how, paragraph three what was found, and paragraph four the significance. The method is now described, which the examined abstract left out almost entirely. It covers the harmonization of four survey rounds, the analysis sample of 14,963 workers, the four models compared, the stratified 70/15/15 split, the use of SHAP, and the substitution of model predictions into the Oaxaca-Blinder decomposition. The length now meets Section 4.2 of the UCU Academic Research Manual, which limits a Master's abstract to 250 words. The examined abstract ran to 334 words. The revised abstract is 244 words, written as continuous paragraphs and mostly in the past tense, as that section also requires. Three figures have been corrected against the Chapter Four tables. The abstract examined reported 14,917 workers where Chapters 3, 4 and 6 all report 14,963; and highest values of 52.4 and 84.9 per cent where Tables 4.4 and 4.5 give 56.6 and 84.4. Those figures had been taken from the Appendix E poster, which shows results from an earlier run of the pipeline. The abstract now reports the ranges from the final tables, 14 to 57 per cent for the gender gap and 68 to 84 per cent for the rural-urban gap. Random Forest's test R-squared of 0.399 is now reported alongside XGBoost's 0.410, so the gap between the two best models is not made to look bigger than it is. The limit to formal paid employment is now stated in the sample description, and the closing sentence says that the unexplained component is a residual, not a measure of discrimination.	Abstract (preliminary pages)

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
3	The research title needs revision so that it is accurate, clear, succinct, and clearly communicates what is being investigated.	Title revised to: “Machine Learning Decomposition of Gender and Rural-Urban Wage Gaps: Evidence from Uganda National Household Surveys”. The examined title named the method and the data source but never said which gaps were being studied. The revised title names both gaps, says what was done (decomposition) instead of only naming the method, and keeps the data source. To be updated on the cover page, in the running header, on the Appendix E poster and in the Appendix H supervision plan heading.	Cover page; running header; Appendix E; Appendix H
4	Explicitly link the reviewed literature to the purpose of the study.	A closing paragraph was added to Section 2.1. It explains that each review section now ends by saying what the work it covers leaves unanswered, and that Section 2.8 matches those points to the objectives, so a reader can trace every objective back to a specific gap in the published work. Closing paragraphs linking the literature to the purpose of the study were added at the end of Sections 2.5, 2.6 and 2.7. Each one names the objective it supports: Section 2.5 and Section 2.6 to Objective 1, Section 2.7 to Objective 2.	Sections 2.1, 2.5, 2.6, 2.7; pp. 10, 16-18
5	Demonstrate how the literature reviewed provides the rationale for conducting this research.	New subsection 2.8.1, “How the Reviewed Literature Provides the Rationale for This Study”, has been added between the synthesis prose and Table 2.3. It lists four findings from the reviewed work and the design choice each one supports: returns to schooling curve upward and differ sharply between groups, which is a reason to compare models that allow more and more flexibility; the German and Swiss results point to something that can be tested, not just observed; the SHAP additivity results make it possible to attribute results from a tree model, and without that the study could not answer its questions; and Mora and Herrera (2025)	Section 2.8.1, p. 18

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
		together with the harmonization literature show that repeated cross-sections can answer a multi-round question as long as the harmonization is documented.	
6	Move beyond simply reporting results to interpreting what they mean in relation to the study and existing knowledge.	Interpretive closing passages have been added at the end of Sections 4.2.1, 4.2.2 and 4.2.3. Each one discusses the result in light of the objective it answers and of named earlier studies, instead of simply repeating the table. Section 4.2.1 discusses the 87.9 per cent retention rate, and how it varies by round, in relation to normal practice for harmonizing surveys after collection. Section 4.2.2 argues that the predictive gain came from allowing nonlinearity rather than from the boosting algorithm, and discusses the drop in importance of age squared in relation to Mincer (1974) and Lemieux (2006). Section 4.2.3 argues that the two gaps are different in kind, that relaxing the shape of the model changes the conclusion about the gender gap instead of just refining it, and that this shift does not explain away the unexplained component.	Sections 4.2.1-4.2.3; pp. 28-36

B. COMMENTS BY INTERNAL EXAMINER (SUPERVISOR)

S/N	COMMENTS BY INTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	No separate written comments were issued by the internal examiner for this submission.	The examiner and panel comments recorded in the tables above and below were discussed with the supervisors, Dr Samuel Kakuba and Eng. Ian Raymond Osolo, and the corrections listed in the “Action Taken” columns were agreed with them before they were made. This row will be replaced if written internal-examiner comments are subsequently received.	Not applicable

C. COMMENTS BY VIVA VOCE PANEL

S/N	COMMENTS BY VIVA VOCE PANEL	ACTION TAKEN	INDICATOR
1	Can the choice of design science as a research methodology be revised, as there is no artefact produced?	Design Science Research has been dropped as the stated research design. Section 3.2 has been rewritten to describe the study as a quantitative, non-experimental, comparative design built on secondary survey microdata, and to set out the two comparisons it supports: between models on the same test set, and between survey rounds using one model setup. A paragraph explains why the label was dropped. The study produces analytical results and a documented, reusable pipeline, but it does not build, deploy or evaluate a system with intended users, and it did not include the demonstration or evaluation stage that Hevner et al. (2004) and Peffers et al. (2007) describe. Former subsection 3.2.1 has been deleted and former 3.2.2 (Pragmatism) renumbered. The pipeline is now reported as a research output in Section 6.7 instead of a designed artefact.	Section 3.2, pp. 20-21; Section 6.7, p. 47
2	How was the explainability aspect explained?	New subsection 3.4.1, “How Explainability Was Implemented”, has been added. It explains that all of the explainability comes from SHAP, and that the study does not claim Random Forest or XGBoost can be read directly. It gives the additivity identity $f(x) = \phi_0 + \sum \phi_k$, names TreeExplainer as the method used, which gives exact values for tree ensembles instead of sampled ones, and lists the four outputs produced: the ranking of mean absolute attribution across all four models, the signed global attribution plot, the dependence plots, and the separate female and male subgroup attributions. It then shows how this links to the decomposition. Averaging the identity within each group and subtracting gives $\text{explained}_k = \text{mean}(\phi_k \mid A) - \text{mean}(\phi_k \mid B)$, the base value cancels out, and adding these up across features gives the explained component exactly. The	Section 3.4.1, p. 25

S/N	COMMENTS BY VIVA VOCE PANEL	ACTION TAKEN	INDICATOR
		feature-by-feature breakdown is therefore the same quantity as the decomposition, read one feature at a time. Two limits are noted: SHAP explains a prediction, not a real-world outcome, and its attributions become unstable when features are correlated, as education, occupation and industry are in this data.	
3	The clarification on the dataset – panel, longitudinal, cross-sectional – should be explained clearly.	New subsection 3.3.1.2, “Structure of the Data: Repeated Cross-Sections, Not a Panel”, has been added. It explains the difference between the three terms, notes that each UNHS round draws a fresh two-stage stratified cluster sample so that no one person appears twice, and calls the resulting structure a pooled repeated cross-section. It then says what that structure allows: comparing population figures across rounds, and fitting one model to all four rounds with round included as a control. It also says what it does not allow, which is any claim about how one person’s wages changed over time. The word “longitudinal” is now used only to mean repeated measurement of the same population, and is replaced by “pooled multi-round dataset” where precision matters: in the abstract, Objective 1, Research Question 1, the Section 2.5 heading, Section 3.2, and Chapters 4 and 6.	Section 3.3.1.2, p. 21; plus Abstract, 1.4.2, 1.5, 2.5, 3.2, 4.1, 6.2
4	How was the under-representation of the female gender in the dataset handled, and how did it affect the results?	New subsection 3.3.4, “Under-Representation of Women, and How It Was Handled”, has been added. It gives the figures: the female share of wage reports ranged from 34.1 per cent in 2019/20 to 38.6 per cent in 2023/24, and the 14,963-record analysis sample contains 9,728 men and 5,235 women, a female share of 35.0 per cent. It shows that the imbalance reflects who actually works for wages in Uganda and is not a fault in the sampling: most working women are in subsistence,	Section 3.3.4, pp. 22-23; Section 3.8, p. 27

S/N	COMMENTS BY VIVA VOCE PANEL	ACTION TAKEN	INDICATOR
		unpaid or self-employed work that reports no monthly cash wage, so a sample that was half female would not describe Uganda's wage earners. Four steps were taken: the UBOS stratified sampling weights were kept; the 70/15/15 split was stratified by sex as well as by round; the Neumark pooled reference does not require equal group sizes; and the 500-observation pooling trigger and 200-observation hard minimum are applied to the smaller subsample, not the cell total. The effect on the results is on precision, not on the size of the estimates. The smallest female cell had 607 observations, so the LASSO fallback was available but never triggered. Four of the sixteen round-level cells needed adjacent-round pooling, all of them gender-by-residence cells for 2012/13 and 2019/20, and these have wider bootstrap intervals. A separate issue, that the sample only includes women who had already found wage work, is cross-referenced to Constraint 6 in Section 3.8.	
5	The study should not over-claim by saying that the wage gap is evidence of gender discrimination. The unexplained component is often associated with discrimination but is not the same thing.	The definition of "Unexplained Component" in Section 1.11 has been rewritten. It now says that the term is not a measure of discrimination, and lists what the residual can include: hours actually worked, job tenure, firm size and productivity, effort, professional networks and reporting error, together with any unequal treatment of otherwise similar workers. A paragraph headed "What the unexplained component is not" was added at the end of Section 5.4 and again in Section 6.3. It says that nothing in the study identifies discrimination, that a residual of any given size could just as easily come from several causes that are not discrimination, and that proving discrimination would need a study designed for that purpose.	Section 1.11, p. 8; 1.8, p. 6; 5.4, p. 41; 6.3, p. 44; Appendix D

S/N	COMMENTS BY VIVA VOCE PANEL	ACTION TAKEN	INDICATOR
		The wording was tightened in Section 1.8 and in Appendices D.2.1 and D.3.1, where the earlier text jumped straight from a large unexplained share to discrimination. The rewritten abstract carries the same caveat.	
6	It would be interesting to see the results of the different sub-groups and the general dataset trained with different techniques.	New subsection 4.2.4, “Subgroup Results Across Estimation Techniques”, has been added, with a new Table 4.8 placing all four subgroup splits side by side on the pooled four-round estimate, each with the gap, the OLS explained share and the ML explained share. The comparison shows that the choice of technique matters little where a gap is driven by differences in worker characteristics, and matters a great deal where it is not: the two techniques are about 22 points apart on the urban/rural split but about 34 points apart, and on opposite sides of zero, on the gender split. Within the gender gap, the linear method explains 11.0 per cent of the gap among urban workers but only 0.5 per cent among rural workers, against 40.4 and 34.7 per cent under the ML method, so the linear method struggles most in exactly the subgroup that carries the study’s most policy-relevant finding. The subsection also says what was not run, and why. LASSO was kept as a fallback but never used, because no cell fell below the 200-observation minimum. Random Forest was not put through the full four-round, four-split, bootstrapped decomposition, on runtime grounds.	Section 4.2.4 and Table 4.8, p. 36
7	Since Random Forest and XGBoost have similar results, why wasn’t Random Forest chosen, since it is more explainable?	New subsection 5.3.3, “Why XGBoost Rather Than Random Forest”, has been added, giving three reasons and one admission. First, the rule for choosing a model was set in Section 3.4 before the comparison was run: whichever model had the lowest error on held-out data would be carried forward. XGBoost was better on all three metrics and on all three splits, with test RMSE of 0.9110 against 0.9194,	Section 5.3.3, p. 39; Section 6.6.1, p. 46

S/N	COMMENTS BY VIVA VOCE PANEL	ACTION TAKEN	INDICATOR
		test MAE of 0.6016 against 0.6059 and test R^2 of 0.410 against 0.399. Changing that rule after seeing the results would have been the weaker choice. Second, the two models are equally easy to interpret here, because the interpretation does not come from the model itself. Random Forest's reputation for transparency applies to a single shallow tree, not to an ensemble of several hundred tuned trees that no reader is going to check rule by rule. Explainability here is supplied by SHAP through TreeExplainer, which is exact for both models and produces identical outputs for each. Third, the main conclusions do not depend on which of the two was chosen. Table 4.3 shows the two models agreeing on education as the leading feature, at 0.3216 and 0.3105, agreeing in demoting age squared, and placing occupation, sex, region and industry within a narrow band. The conclusion that the gain came from allowing nonlinearity, and not from one particular algorithm, rests on that agreement. The subsection says openly that this reasoning was not checked by re-running the decomposition with Random Forest, and that check has been added to Section 6.6.1 as further work.	

D. ADDITIONAL CORRECTIONS MADE ON REVIEW

The following were not raised by the examiners or the panel. They were found while implementing the corrections above, checked against the UCU Academic Research Manual (revised 2018), and corrected in the same revision.

- Appendix D described five survey rounds covering 2009/10 to 2023/24 and said twice that the data covered fifteen years of change. The study uses four rounds covering twelve years, 2012/13 to 2023/24. Corrected in three places.
- Appendix A recorded a stratified 80/20 train/test split. Section 3.4 and Chapter Four both report a 70/15/15 train, validation and test split stratified by round and sex. Corrected.
- Appendix C cited Kagundu and Pavlova as a Journal of African Economies article with an impact factor. The bibliography correctly records it as Working Paper 07-26, Andrew Young School of Policy Studies, Georgia State University. Corrected, together

with the Type and IF columns. Cengiz et al. was dated 2021 here and 2022 in the bibliography; set to 2022.

- The Ethical Generative AI Usage Declaration was appearing in the Table of Contents as “Declaration”, so two different pages showed up under the same entry. Corrected to the page’s own title.
- The Dedication heading had been commented out while its contents line and input were left active, so the Table of Contents pointed to a page with no heading on it. The heading has been restored.
- The Abstract has been moved to sit immediately after the title page, the position given in Section 4.1.1 of the Research Manual. It previously sat fourth among the preliminary pages.
- Sections 4.7, 4.8 and 4.9 of the Research Manual require a List of Tables, a List of Figures and a List of Appendices. Version 03 had none of the three. To be added to the preliminary pages.
- The gap is spelled three ways across the dissertation: urban-rural in Chapter One, rural/urban in Chapters Four to Six, and Rural/urban in Tables 4.5 and 4.7. This is being standardized to rural-urban throughout, to match the revised title.
- The Appendix E poster carries a results series from an earlier run of the pipeline, which does not match Tables 4.4 and 4.5. It will be regenerated from the final tables, together with the revised title.

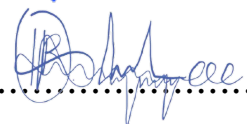
Candidate’s Name: Bob Robert Tumushiime
Date: 15th September 2026

Signature: 

Supervisor’s Name: Dr Samuel Kakuba
Date: 22/09/2026

Signature: 

Co-Supervisor’s Name: Eng. Ian Raymond Osolo
Date: 22/09/2026

Signature: 

NB: The Post Viva compliance form captures all corrections recommended by the internal examiner (supervisor), the external examiner and the viva panel. Page numbers in the Indicator column refer to version 03 of the dissertation and will shift in the revised copy; final page numbers will be inserted before binding.