**SUPPLEMENTARY MATERIAL**

This supplementary material is to submit to Ageing & Society journal for review on the manuscript, entitled “Socioeconomic-related health inequality in non-communicable diseases among older people in Viet Nam”. This material provides results of model specification tests and comparison among selected models.

**Model specification tests**

We used Pregibon’s link test to examine model specification. In brief, the logic of that test is to regress the dependent variable on its predicted values and square of its predicted values. Such test provides us information on whether important variables correlated with high-order terms are omitted or not. More specifically, Pregibon’s link test assesses whether the coefficients on the squared terms are significant difference from zero. Detailed information of Pregibon’s link test has been presented elsewhere (Deb, Norton, & Manning 2017[[1]](#footnote-1); Pregibon 1980[[2]](#footnote-2)). After running OLS (for a continuous outcome) or probit regression (for a binary outcome), we used postestimation linktest in Stata to execute the specification tests. If a *p*-value of the square term is statistically insignificant, then ones can conclude that models are well-specified (Deb, Norton, & Manning 2017). A *p*-value that was less than 0.05 was regarded as statistically significant.

* 1. Probit models using having at least one NCD as a dependent variable.
	2. Urban areas



* 1. Rural areas



It can be seen that a *p*-value of the square term (\_hatsq) was statistically insignificant in models for both rural and urban areas, implying that models used were well-specified.

* 1. Probit models using multi-morbidity as a dependent variable

2.1) Urban areas



2.2) Rural areas



In the case where a dependent variable was measured as multi-morbidity, model for rural areas was specified but model for urban was unspecified (a *p*-value of the square term <0.05).

* 1. Linear models

3.1) Urban areas



3.2) Rural areas



In the case where a dependent was a continuous variable, model for urban was specified but a *p*-value of the square term in model for rural areas was quite sensitive (0.05). We suspected that Stata may round up the *p*-value to the nearest second decimal place, thus we tried to compute link test manually and results of such test were shown below:



Our suspicion was correct that Stata rounded up the *p*-value. Here, it can be seen that a *p*-value of the square term for rural areas was less than 0.05, thus we sufficiently concluded that model for rural areas was unspecified.

**Model comparison using log likelihood and information criteria (AIC and BIC)**

In this exercise, we used log likelihood, AIC, and BIC produced by each model to compare amongst linear models and probit models (either using having at least one NCD or multi-morbidity as a dependent variable). Models with higher values in log likelihood and lower values in AIC and BIC were preferred. The results showed strong evidence in favour of probit models that used having at least one NCD as a dependent variable. Thus, we concluded that those models were the best-fitting ones.

|  |  |  |
| --- | --- | --- |
| **Model** | **Urban areas** | **Rural areas** |
| **Log likelihood** | **AIC** | **BIC** | **Log likelihood** | **AIC** | **BIC** |
| Linear regression | -1244.747 | 2549.495 | 2686.155 | -3356.156 | 6772.312 | 6940.022 |
| Probit models – Having at least one NCD | -354.0723a | 768.1446a | 904.8053a | -1109.18a | 2278.359a | 2446.070a |
| Probit models – Multi-morbidity | -444.3727 | 948.7455 | 1085.406 | -1296.012 | 2652.024 | 2819.734 |

*Note: apreferred models*

1. Deb, P., Norton, E. C., Manning, W. G. (2017). *Health econometrics using Stata*. College Station, TX: Stata Press. [↑](#footnote-ref-1)
2. Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of Royal Statistical Society, 29*(1), 15-23 [↑](#footnote-ref-2)