Variable Scaling and Joint Hypothesis Testing in PPML*

Anton Yang† Russell Hillberry‡

Yale University Purdue University

September 21, 2026

Abstract

The enormous gravity-model-of-trade literature has illustrated important advantages of estimation via Poisson Pseudo Maximum Likelihood (PPML). That literature's prioritization of parameter estimation over hypothesis testing has left questions of testing in the PPML framework underexplored. In this paper we show analytically that scaling the dependent variable can affect the outcome of some joint hypothesis tests, but not others. Likelihood-ratio, model-based Wald, and model-based Lagrange Multiplier test statistics depend on scale, and therefore do not support scale-invariant inference. Wald and Lagrange Multiplier tests constructed with heteroskedasticity-robust sandwich adjustments are invariant to scale. We illustrate these points empirically with an application from the literature on Revealed Comparative Advantage.

Keywords: Poisson Pseudo Maximum Likelihood; Variable scaling; Revealed comparative advantage; Wald test; Joint hypothesis testing

JEL classification: C12; C13; C81; F14

*Declarations of interest: None.

†A307, 28 Hillhouse Avenue, New Haven, CT 06511. Email: anton.yang@yale.edu.

‡*Corresponding author*: 692 Krannert Building, 403 Mitch Daniels Blvd., West Lafayette, IN 47907-2053. Email: rhillber@purdue.edu.

1 Introduction

The enormous gravity-model-of-trade literature has demonstrated a number of important benefits of estimation via Poisson Pseudo Maximum Likelihood (PPML).[1] The gravity literature's focus on parameter estimation has left some important questions related to hypothesis testing underexplored. In particular, the question of how best to evaluate joint hypotheses in the PPML framework has not yet been well answered. Since PPML is likely to be a useful tool in other literatures where joint hypothesis testing is important, a guide to conducting such tests should be useful to other researchers.

In this paper we consider the effect of a researcher's data scaling choices on the robustness of different approaches to testing joint hypotheses in the PPML framework. Many potential applications of the framework study outcome variables without a well-identified unit. In the case of the gravity model and other trade applications the dependent variables are monetary values, which researchers often scale for purposes of convenience. In this paper we show that data scaling choices can affect the outcome of seemingly plausible joint hypothesis tests. This is unfortunate because PPML's attractive qualities apply to many questions where the data are scalable and tests of joint hypotheses are interesting. Our purpose is to guide researchers toward estimating strategies that produce tests of joint hypotheses that are invariant to data scaling choices.

We use analytical methods to demonstrate the issues. We begin with the (pseudo) likelihood-ratio statistic constructed from the PPML objective. In our application, as in many PPML applications, the dependent variable is a continuous monetary value rather than a count. The PPML objective is therefore a pseudo-likelihood criterion rather than a proper likelihood for the observed data, and the resulting statistic does not have the usual

[1] Among the most important benefits relative to log-linear OLS are a) that PPML allows the estimation of elasticities when zero observations enter on the left hand side (Santos Silva and Tenreyro, 2011), b) that the model maps cleanly onto structural theories of gravity with a constant elasticity of substitution between regions' factor supplies (Anderson and Yotov, 2012; Fally, 2015), c) that the inclusion of fixed effects naturally imposes relevant adding up conditions (Fally, 2015), and d) that the estimator weights observations by the scale of the predicted values, which is useful in some contexts (Correia, Guimarães and Zylkin, 2020).

likelihood-ratio interpretation (Green and Santos Silva, 2025). We nevertheless begin with this statistic because its sensitivity to data scaling provides a particularly transparent illustration of the problem. We next show that the Wald and Lagrange Multiplier tests that use the model-based information matrix are also sensitive to scaling. By contrast, a Wald test constructed from a Huber–White sandwich covariance matrix and an LM test constructed from the empirical covariance matrix of the score contributions are invariant to scale. In this sense, the heteroskedasticity-robust adjustment solves the scaling problem for Wald and LM inference. While both robust Wald and robust LM methods are valid, we view the Wald test as the more practical choice.

In order to illustrate the lessons of our analytical work, we use an application from the literature on Revealed Comparative Advantage (RCA). A canonical paper in this literature is Nunn (2007), which uses log-linear regressions to show that countries with stronger legal institutions have RCA in sectors using inputs that require contract-intensive market transactions. Hillberry, Jimenez and Karabay (2025) note that the PPML estimator's implicit adding up condition and its consistent dollar-weighting properties are conceptually useful for studies of RCA. We revisit Nunn's estimates using a PPML specification. The key coefficient of interest in Nunn (2007) remains positive and statistically significant when the model is estimated with PPML.

One of Nunn's robustness checks identifies an opportunity to showcase our lessons about joint hypothesis testing in the PPML framework. We show that the coefficient estimates associated with two interactions that Nunn considers are jointly statistically significant, even though the individual coefficients are not statistically different than zero. We also validate our analytical results about each strategy's robustness (or lack thereof) to scaling the independent variable.

The remainder of the paper is organized as follows. Section 2 reviews the PPML estimator. Sections 3–5 demonstrate analytical results for the Likelihood Ratio test, the Wald test, and the Lagrange Multiplier test, respectively. Mathematical derivations for sections 4 and

5 are relegated to appendices. In section 6 we show the empirical application to RCA, examining joint restrictions in PPML analogues of Nunn's OLS estimates and their sensitivity to scaling. Section 7 concludes.

2 The PPML Estimator

Consider the estimation specification suggested by Santos Silva and Tenreyro (2006a):

$$y_i = \exp(x_i'\beta) + \epsilon_i, \tag{1}$$

where y_i, $i = 1, .., n$, can be observed from data, x_i is a vector of exogenous variables, β a vector of associated parameters, and ϵ_i an error term with $E[\epsilon_i|x_i] = 0$. Following Gourieroux, Monfort and Trognon (1984*a*,*b*), choose β to maximize the Poisson pseudo-log-likelihood criterion:

$$L(\beta) = K - \sum_i^n \exp(x_i'\beta) + \sum_i^n y_i(x_i'\beta), \tag{2}$$

where K is a constant term.

Our analytic exercises will use this general framework, the our lessons can be expressed by a simplified version with two right hand side variables.

$$y_i = \exp[\beta_0 + \beta_1 x_i + \beta_2 z_i] + \epsilon_i, \tag{3}$$

where y_i represents a scalable independent variable (e.g. international trade), x_i and z_i represent dependent variables in the model, and ϵ_i is a mean zero error term.

Combining (2) and (3), the (pseudo) log-likelihood function is given by:

$$
\begin{aligned}
L(\beta) = & K - \sum_{i}^{n} \sum_{j}^{n} \exp[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i)] \\
& + \sum_{i}^{n} \sum_{j}^{n} y_{ij} [\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i)].
\end{aligned} \tag{4}
$$

3 Scale and Testing using Likelihood Ratios

Given the prominence of the word "likelihood" in the name of the estimating framework, one might be tempted to employ a likelihood-ratio test when evaluating joint restrictions. For continuous outcomes such as trade values, however, the PPML objective is a pseudo-likelihood criterion rather than the log likelihood of a fully specified probability model. A conventional likelihood-ratio interpretation is therefore not appropriate (Green and Santos Silva, 2025). We nevertheless examine the statistic because it provides a transparent illustration of how the units of the dependent variable can affect inference.[2] The statistic is also available in widely used PPML software.[3] We walk the reader through the effect of scaling on the likelihood statistic and likelihood ratio test as an illustration of the general problem we discuss here.

Consider two models: a 'small' parsimonious model, and a 'large' model with more parameters. We wish to consider a hypothesis test of the form:

Hypothesis $\mathbf{H_0}$: *the 'small' model is more consistent with the data.*

Hypothesis $\mathbf{H_a}$: *the 'large' model is more consistent with the data.*

[2]For Poisson count data, the full log-likelihood includes the normalizing term $-\log(y_i!)$, which replaces the 'K' term in (2). Count data have a natural unit, so are typically not subject to arbitrary multiplicative rescaling that changes the statistical experiment. In PPML applications with continuous outcomes such as trade values, however, rescaling may represent only a change in reporting units. The Poisson objective in this case is a pseudo-likelihood criterion, and an LR statistic constructed from its maximized value is not invariant to that change. Including the $-\log(y_i!)$ term would not itself eliminate the problem because it cancels between the restricted and unrestricted models.

[3]Correia, Guimarães and Zylkin (2020)'s `ppmlhdfe` command, for example, reports the log-pseudo likelihood statistic.

The test statistic for an LR test of $\mathbf{H_0}$ would appear as:

$$\Lambda = -2[L^{H_0}(\beta) - L^{H_a}(\beta')], \tag{5}$$

where β and β' are the parameter vectors associated with the 'small' and 'large' models, respectively. Asymptotically, Λ is distributed χ^2 with k degrees of freedom, where k is the reduction in the number of parameters when moving from the 'large' to the 'small' model. Given a critical value c for this distribution, the null hypothesis would be rejected if $\Lambda > c$.[4]

3.1 Implications of Scaling for Likelihood

Let the log-likelihood function for the 'small' model take the form:

$$L^{H_0}(\beta) = -\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + \sum_i^n y_i[\beta_0 + \beta_1 \log(x_i)], \tag{6}$$

where β_0 is a constant term, β_1 is the coefficient on $\log x_i$.[5]

In order to show that the scaling of y_i affects Λ, we introduce an arbitrary scalar $S \neq 1$ that we apply to y_i. The associated log-likelihood function is:

$$\widetilde{L^{H_0}}(\tilde{\beta}) = -\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + \sum_i^n S y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] \tag{7}$$

where $\widetilde{L^{H_0}}(\tilde{\beta})$ is the log-likelihood involved with the scaling factor S, and $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are the model's parameters given the scaling. Taking the first-order conditions with respect to β terms in Equations (6) and (7) and solving the associated system of equations reveals relationships between the parameters before and after the scaling: $\beta_0 = \tilde{\beta}_0 - \log(S)$ and $\beta_1 = \tilde{\beta}_1$. Intuitively, scaling affects the size of the constant term, but other coefficients in the model are unaffected. However, the log-likelihoods before and after the scaling of y_i are

[4] See Gouriéroux, Holly and Monfort (1982).

[5] We follow the standard PPML-gravity literature and suppress the parameter K, which cancels when constructing LR statistic in equation (5).

not equivalent:

$$\begin{aligned}\widetilde{L^{H_0}}(\tilde{\beta}) =& -\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + S\sum_i^n y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] \\ =& -S\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + S\sum_i^n y_i[\beta_0 + \log S + \beta_1 \log(x_i)] \\ \neq& -\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + \sum_i^n y_i[\beta_0 + \beta_1 \log(x_i)] \\ =& L^{H_0}(\beta).\end{aligned} \tag{8}$$

The relationship only holds with equality when $S = 1$.

3.2 Implications of Scaling for Testing

Now shift to the large model by introducing the z_i variable. The associated log-likelihood function is:

$$L^{H_a}(\beta') = -\sum_i^n \exp[\beta'_0 + \beta'_1 \log(x_i) + \beta'_2 \log z_i] + \sum_i^n y_i[\beta'_0 + \beta'_1 \log(x_i) + \beta'_2 \log z_i], \tag{9}$$

with $'$ indicating parameters of the large model. The log-likelihood of the large model with scaled data appears as:

$$\widetilde{L^{H_a}}(\tilde{\beta}') = -\sum_i^n \exp[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)] + S\sum_i^n y_i[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)]. \tag{10}$$

It can be shown that the parameters in the scaled and unscaled large models have the relationships: $\beta'_0 = \tilde{\beta}'_0 - \log(S)$, $\beta'_1 = \tilde{\beta}'_1$ and $\beta'_2 = \tilde{\beta}'_2$.

The difference between the log-likelihoods using the scaled data is given by:

$$\widetilde{L^{H_0}}(\tilde{\beta}) - \widetilde{L^{H_a}}(\tilde{\beta}') = -\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + S\sum_i^n y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)]$$
$$+ \sum_i^n \exp[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)] - S\sum_i^n y_i[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)]. \tag{11}$$

Dividing both sides by S in equation (11) and substituting the derived parametric relationships into the test statistics produces:

$$\begin{aligned}\frac{\widetilde{\Lambda}(\tilde{\beta})}{S} = \frac{-2\left[\widetilde{L^{H_0}}(\tilde{\beta}) - \widetilde{L^{H_a}}(\tilde{\beta}')\right]}{S} = -2\Bigg\{ &-\frac{1}{S}\sum_i^n \exp[\beta_0 + \log(S) + \beta_1 \log(x_i)] \\ &+ \sum_i^n y_i[\beta_0 + \log(S) + \beta_1 \log(x_i)] \\ &+ \frac{1}{S}\sum_i^n \exp[\beta'_0 + \log(S) + \beta'_1 \log(x_i) + \beta'_2 \log(z_i)] \\ &- \sum_i^n y_i[\beta'_0 + \log S + \beta'_1 \log(x_i) + \beta'_2 \log(z_i)]\Bigg\}.\end{aligned} \tag{12}$$

The right-hand side of equation (12) is equal to the test statistic with unscaled data: $\Lambda(\beta) = -2\Big[L^{H_0}(\beta) - L^{H_a}(\beta')\Big]$. The lesson is that $\Lambda(\beta)$ varies inversely with the scale applied to the y variable. The critical value of the χ^2 distribution is unchanged, but the test statistic changes with S, polluting statistical inference.

4 Wald Test

We next turn our attention to the Wald test, where we observe that the effects of scaling depend upon the errors that are used to calculate the Fisher information matrix. We show analytically that test statistics calculated with model-based asymptotic standard errors depend on data scaling, while test statistics calculated with robust standard errors do

not.[6] To simplify the exposition, we present the full analytical derivations in Appendix A, where we use the 'small' model and a restriction on β_1 to show that the non-robust Wald statistic scales with the dependent variable, while the robust Wald statistic is invariant to such scaling. The invariance follows from the sandwich construction: scaling multiplies the information matrix by S and the empirical covariance of the score contributions by S^2, so the scale factors cancel. These results extend directly to multivariate settings with multiple parameters, as shown in Appendix B.

5 Lagrange Multiplier (LM) Test

We then turn to the LM test, where we examine how scaling affects inference based on the score and information matrix evaluated under the null. We show that the conventional LM statistic depends on the scale of the dependent variable, while a heteroskedasticity-robust LM statistic is invariant to such scaling. These results extend to more general specifications with additional parameters. As in the Wald case, Appendix B establishes the scaling properties of the information matrices that support this extension to LM and related test statistics. We present the full analytical derivations in Appendix C, where we consider the 'large' model and a restriction on β_2, and show that the non-robust LM statistic scales proportionally with the dependent variable, whereas the robust LM statistic does not.

The conventional LM statistic can also be adjusted mechanically.[7] Since $\text{LM}(S) = S\text{LM}(1)$, dividing the statistic by the known scale factor returns the statistic under a chosen baseline normalization. This adjusted LM statistic is useful for illustrating the scaling result, but it is less attractive as a general recommendation because it requires the researcher to identify the relevant baseline scale and continues to rely on the model-based variance assumption. The robust LM statistic avoids both limitations, but requires estimation under

[6] The "model-based" qualifier on the asymptotic standard errors is formal language acknowledging that the errors depend upon the model specification as well as upon the estimated parameters. In the interest of brevity, we omit the "model-based" adjective hereafter.

[7] We thank an anonymous referee for raising this point.

the null and construction of the nuisance-adjusted score covariance matrix. By contrast, the robust Wald statistic can be calculated directly from the unrestricted PPML estimates and the robust covariance matrix routinely reported by standard software. For routine applied work, we therefore recommend the robust Wald test.

The empirical application in Section 6 confirms these results. As shown in Table 2, the model-based LM statistic changes in direct proportion to the scale of export values, while the adjusted LM, robust LM, and robust Wald statistics are unchanged.

6 Revealed Comparative Advantage and Joint Testing

Our theoretical and empirical results thus far show that LR tests and some Wald or LM tests are sensitive to arbitrary choices of scale for the dependent variable in PPML estimation. A natural place to seek an economically meaningful joint test is the literature on the determinants of comparative advantage, where blocks of interactions have clear economic content.[8] As an example of our analytical lessons, we use data from Nunn (2007) to replicate his results and re-estimate with PPML. Nunn's purpose is to test the proposition that countries with better legal institutions have RCA in sectors that use contract-intensive inputs. We use parameters from one of Nunn's robustness checks to illustrate our analytical lessons from above.

6.1 Data and Empirical Strategy

We use Nunn's replication files to construct a country–industry panel of exports in 1997. The data cover 146 countries and approximately 220 three-digit SITC industries. For each country c and industry i we observe exports Y_{ci}, Nunn's contract-intensity measure z_i^{rs1},

[8]Given PPML's heavy use in the gravity model of trade literature, the gravity model might seem like a better setting for an illustration. Unfortunately, the gravity literature offers few useful applications where testing of joint hypotheses is an informative exercise. One exception where which joint tests can be useful are structural estimation exercises. Yang (2021) estimates a gravity model with constant-difference-of-elasticities preferences and jointly evaluates restrictions implied by the preference specification using a robust Wald test.

factor-intensity measures, and country characteristics such as judicial quality, per capita income, human-capital abundance, capital abundance, and financial development. The PPML specification follows the treatment of RCA in Hillberry, Jimenez and Karabay (2025), who suggest estimating PPML with raw export values and country and industry fixed effects rather than placing a share variable on the left-hand side. This approach preserves the adding-up structure of trade flows and avoids biases associated with log-linear transformations and missing values when zero trade flows are observed. We estimate

$$Y_{ci} = \exp\left[\alpha_c + \alpha_i + V_c X_i \Upsilon\right] + \varepsilon_{ci}, \tag{13}$$

where α_c and α_i are country and industry fixed effects, V_c are country characteristics, X_i are industry characteristics, Υ is a vector of parameters to be estimated and ε_{ci} a mean zero error term. This specification is useful in the RCA setting because the fixed effects absorb country export totals and world industry export totals, so the identifying variation comes from interactions of country characteristics with industry characteristics. It also allows the PPML's implicit observation weights to align with fitted export values, a key reason the setup is attractive in the RCA context.[9] Unlike Nunn's original log-linear OLS, the PPML estimates use observations with zero exports and do not place a logarithm or a share on the left-hand side.

We begin by replicating the central OLS results in Nunn (2007). We then re-estimate with the PPML estimator. We focus on column 3 and 5 in Nunn's main results (reported in Table IV there). Column 3 of that table reports results from a simple three-variable specification that includes controls for interactions between relative factor endowments and intensities. Column 5 is a richer specification that Nunn includes to check robustness of his main results. We report both Nunn's original log-linear OLS results and our PPML results

[9]PPML weights observations in proportion to their predicted values in estimation (Correia, Guimarães and Zylkin, 2020). This property is neutral across variation in sectors i and countries c. For a given proportional size ratio of two fitted trade flows (either within a sector or within two countries) the relative weights in the estimation will be equi-proportional in both cases.

in Table 1, with the OLS results appearing in columns 1 and 2, and the PPML results appearing in columns 3 and 4.

Table 1: Determinants of Comparative Advantage: OLS and PPML Estimates

	OLS		PPML	
	Nunn col. (3)	Nunn col. (5)	Col. (3) analogue	Col. (5) analogue
Judicial quality × contract intensity	0.326**	0.296**	3.314**	3.383**
	(0.023)	(0.024)	(0.769)	(0.751)
Skill intensity × human-capital abundance	0.085**	0.063**	0.465*	0.331
	(0.017)	(0.017)	(0.214)	(0.246)
Capital intensity × capital abundance	0.105**	0.074	0.196*	0.230*
	(0.031)	(0.041)	(0.087)	(0.097)
Value added × log income per capita		-0.137*		0.619
		(0.067)		(0.368)
Intra-industry trade × log income per capita		0.546**		0.752**
		(0.056)		(0.209)
TFP growth × log income per capita		-0.010		-1.433
		(0.049)		(2.053)
Capital intensity × financial development		0.021		-0.024
		(0.018)		(0.102)
Input variety × log income per capita		0.522**		0.265
		(0.103)		(0.498)
Country fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
R^2 / pseudo-R^2	0.760	0.760	0.712	0.714
Observations	10,976	10,816	12,740	12,376

The OLS columns reproduce columns (3) and (5) of Nunn's Table IV and report his published standardized beta coefficients and robust standard errors. We reproduce these estimates using Nunn's replication data. The PPML columns use export values in levels, retain zero exports, and include country and industry fixed effects. Their coefficients enter the exponential conditional-mean specification and are therefore not directly comparable in magnitude with the standardized OLS coefficients. Robust standard errors are in parentheses. The PPML fit statistic is Efron's pseudo-R^2. * and ** denote significance at the 5 and 1 percent levels.

Columns 1 and 3 in Table 1 report the simpler specification from Nunn's column 3. The OLS estimates reproduce Nunn's published results. Because PPML retains zero export observations, the sample rises from 10,976 to 12,740 observations. The interaction of judicial quality with contract intensity remains positive and statistically significant, and both factor-proportions interactions are also significant at the 5 percent level under PPML. Nunn's primary result is verified by estimation via PPML rather than log-linear OLS.

Columns 2 and 4 report the richer specification from Nunn's column 5. The judicial-

quality interaction remains positive and statistically significant under PPML, as do the capital-endowment and intra-industry-trade interactions. The estimator matters more for the income-related controls. In the OLS specification, value added interacted with log income per capita is significant at the 5 percent level and input variety interacted with log income per capita is significant at the 1 percent level. Under PPML, both coefficients are positive but individually insignificant at the 5 percent level, with robust p-values of 0.093 and 0.594, respectively.

6.2 Testing Joint Restrictions

The application we wish to highlight is the effect of data scaling on joint hypothesis testing of the coefficients estimated in columns 2 (OLS) and 4 (PPML). Concerned that the estimated effects associated with the judicial quality interaction may represent, instead, effects operating through other channels, Nunn considers a range of sector-country interactions as control variables. He interacts, respectively, sector value added, an intra-industry trade index, TFP growth, and an index of input variety with countries' per capita income, as well as an interaction of sector-level capital per worker with countries' (logged) ratio of credit to GDP.

Since our purpose in this replication is to demonstrate the role of scaling in joint hypothesis tests, we focus on two interactions that share the same country characteristic. Both value added and input variety are interacted with log per capita income in Nunn's column 5 specification. In the PPML analogue (column 4 of Table 1), neither coefficient is individually significant at the 5 percent level. We test

$$H_0 : \ \beta_{va\times y} = \beta_{m\times y} = 0, \tag{14}$$

where $\beta_{va\times y}$ is the coefficient on value added $\times$ log income per capita and $\beta_{m\times y}$ is the coefficient on input variety $\times$ log income per capita. The restriction has a natural joint

interpretation: it asks whether these two income-related determinants of export composition have explanatory power as a group.

Table 2: Joint Tests of the Income Interactions under Alternative Scales

Scale S	LR	Wald	Robust Wald	LM	Adjusted LM	Robust LM
0.001	0.007 (0.996)	0.007 (0.996)	4.973 (0.083)	0.007 (0.996)	7.414 (0.025)	5.005 (0.082)
1	7.302 (0.026)	7.384 (0.025)	4.973 (0.083)	7.414 (0.025)	7.414 (0.025)	5.005 (0.082)
1,000	7,302.373 (< 0.001)	7,383.667 (< 0.001)	4.973 (0.083)	7,413.743 (< 0.001)	7.414 (0.025)	5.005 (0.082)

Each entry reports a $\chi^2(2)$ statistic with its p-value in parentheses. The null is that the coefficients on value added $\times$ log income per capita and input variety $\times$ log income per capita are jointly zero in the PPML analogue of Nunn's Table IV, column (5). Export values are first multiplied by 10^{-6} for numerical normalization, and S is a multiplicative rescaling relative to that baseline. Wald and LM denote model-based tests. Adjusted LM is $\mathrm{LM}(S)/S$. Robust Wald uses the Huber–White sandwich covariance matrix, while robust LM uses the empirical covariance matrix of the nuisance-adjusted score contributions.

Table 2 confirms the analytical results. The LR, model-based Wald, and model-based LM statistics move in direct proportion to S. At $S = 0.001$, none of these tests rejects; at the baseline scale ($S = 1$), each rejects at the 5 percent level; and at $S = 1{,}000$, each produces an overwhelming rejection. The adjusted LM statistic, defined as $\mathrm{LM}(S)/S$, remains 7.414 with a p-value of 0.025 at every scale. The robust Wald and robust LM statistics are also invariant: the robust Wald statistic is 4.973 ($p = 0.083$), and the robust LM statistic is 5.005 ($p = 0.082$), at each value of S. Thus, the two income interactions are individually insignificant at the 5 percent level, but the robust joint tests reject their joint exclusion at the 10 percent level. The exercise also distinguishes two ways to remove the scaling problem. The adjusted LM test restores a chosen baseline normalization, while the robust Wald and robust LM tests are invariant without an explicit scale correction.

7 Conclusion

PPML estimation offers a number of useful advantages over log-linear OLS, and these advantages have led it to be widely used to estimate parameters that are important in the gravity-model-of-trade literature. Because the focus of the gravity literature is parameter estimation rather than hypothesis testing, the literature has offered relatively little guidance

on best practices for conducting joint hypothesis tests. PPML estimation is potentially useful outside the context of the gravity model, and other literatures may be more likely to find it important to test joint hypotheses.

In this paper we investigate the conditions under which data scaling choices affect hypothesis testing of parameter restrictions imposed on estimates from a PPML model. Tests of individual parameter restrictions can be evaluated with t-tests, so we focus on tests that are capable of evaluating joint restrictions. We use both analytical exercises and empirical examples to demonstrate our results.

Our study considers five kinds of hypothesis tests. We show that data scaling is fatal to those that rely on estimated standard errors that are not robustly estimated. Outcomes of the likelihood ratio test, and versions of the Wald and Lagrange Multiplier tests that do not use robustly estimated residuals depend on the scale of the dependent variable. Wald and Lagrange Multiplier tests that are corrected to account for heteroskedasticity are invariant to data scale. Both robust procedures are valid, but for routine applied work we recommend the robust Wald test because it can be implemented from unrestricted PPML estimates and the robust covariance matrix returned by standard software.

We offer an empirical application of our insights using a canonical paper from the RCA literature. The RCA paper reports log-linear OLS estimates; we estimate the model with PPML instead. While the main result of that paper is robust to the change in estimating strategy, several of the coefficients on control variables change sign and/or lose statistical significance when we estimate with PPML. To demonstrate that the lessons from our analytical exercises hold up in practice, we test a joint hypothesis that two of the interactions entering as control variables are jointly significant. Our results show that the two tests we identified as robust to data scaling issues produce test-statistics and p-values that are robust to scaling of the dependent variable. The other three exercises do not support robust inference. Our exercises can be useful for those who wish to capture the benefits of PPML estimation for empirical questions in which joint hypothesis tests are relevant.

References

Anderson, James E, and Yoto V Yotov. 2012. "Gold Standard Gravity." National Bureau of Economic Research Working Paper 17835.

Cayley, Arthur. 1858. "A Memoir on the Theory of Matrices." *Philosophical Transactions of the Royal Society of London*, 148: 17–37.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson Estimation with High-dimensional Fixed Effects." *The Stata Journal*, 20(1): 95–115.

Fally, Thibault. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics*, 97: 76–85.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984*a*. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica*, 52(3): 701–720.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984*b*. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Gouriéroux, Christian, Alberto Holly, and Alain Monfort. 1982. "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." *Econometrica*, 50(1): 63–80.

Greene, William H. 2012. *Econometric Analysis.* . 7th ed., Prentice Hall.

Green, Nick, and JMC Santos Silva. 2025. "A cautionary note on goodness-of-fit statistics for models estimated by pseudo maximum likelihood." *The Stata Journal*, 25(2): 458–465.

Hillberry, Russell, Manuel Jimenez, and Bilgehan Karabay. 2025. "Geography, Transport and the Composition of Trade." *mimeo*.

Huber, Peter J. 1967. "Under Nonstandard Conditions." Berkeley, CA, USA:University of California Press.

Maas, Cora JM, and Joop J Hox. 2004. "The Influence of Violations of Assumptions on Multilevel Parameter Estimates and their Standard Errors." *Computational Statistics & Data Analysis*, 46(3): 427–440.

Nunn, Nathan. 2007. "Relationship-specificity, incomplete contracts, and the pattern of trade." *The quarterly journal of economics*, 122(2): 569–600.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006a. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Santos Silva, J. M. C., and Silvana Tenreyro. 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator." *Economics Letters*, 112(2): 220–222.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica: Journal of the Econometric Society*, 1–25.

Yang, Anton C. 2021. "Structural Estimation of a Gravity Model of Trade with the Constant-Difference-of-Elasticities Preferences." *mimeo*.

Appendix

A Analytical Derivations for Wald Test

A.1 Non-Robust Asymptotic Standard Errors

To calculate the Wald test statistic with asymptotic standard errors we refer the reader to the unscaled 'small' model in equation (6). The associated Hessian matrix $\boldsymbol{H}(\beta)$ is given by:

$$\boldsymbol{H}(\beta) = \frac{\partial^2 L^{H_0}(\beta)}{\partial\beta \cdot \partial\beta^T} = -\sum_{i=1}^{n} \begin{bmatrix} \exp(\beta_0 + \beta_1 \log(x_i)) & \log(x_i)\exp(\beta_0 + \beta_1 \log(x_i)) \\ \log(x_i)\exp(\beta_0 + \beta_1 \log(x_i)) & (\log(x_i))^2 \exp(\beta_0 + \beta_1 \log(x_i)) \end{bmatrix}. \tag{A.1}$$

The Fisher information matrix $\boldsymbol{I^F}(\beta)$ is the negative of $\boldsymbol{H}(\beta)$. The asymptotic variance-covariance matrix is defined as the inverse of $\boldsymbol{I^F}(\beta)$. Wald test statistics depend on the variance-covariance matrix. Our central result in this section is that scaling the y_i data produces an asymptotic variance-covariance matrix that is a proportional scaling of the asymptotic variance-covariance matrix associated with the unscaled model. We demonstrate this relationship in Theorem 1.

Theorem 1 *Given a twice-differentiable Poisson Pseudo-Maximum Likelihood function in equation (6) with dependent variable y_i for $i = 1, \ldots, n$ and independent variable $\boldsymbol{x}_i \in \mathbb{R}^n$ in the model matrix $\boldsymbol{X} \in \mathbb{R}^{n\times(p+1)}$, and residuals $\boldsymbol{\epsilon} = (y_1 - \exp(x_1'\beta), \ldots, y_n - \exp(x_n'\beta))^\top \in \mathbb{R}^n$. Assume that the Hessian matrix of the function with respect to model parameters $\boldsymbol{\mathcal{B}} = \{\beta_0; \beta_1, ..., \beta_n\}^T$ in the space $\Gamma \subseteq \mathbb{R}^n$ is invertible (nonsingular). If the dependent variable y_i for $i = 1, \ldots, n$ is scaled by a factor $S \neq 1$, then the relationship between the asymptotic variance-covariance matrix $\boldsymbol{V}^{asym}$ of the unscaled model and the asymptotic variance-covariance matrix $\widetilde{\boldsymbol{V}}^{asym}$ of the scaled model must be $\boldsymbol{V}^{asym} = S\widetilde{\boldsymbol{V}}^{asym}$.*

Proof. The asymptotic variance-covariance matrix $\boldsymbol{V}^{\text{asym}}$ stated in Theorem 1 is

$$\boldsymbol{V}^{\text{asym}} = \left[\boldsymbol{I}^F(\beta)\right]^{-1} = \left[-\boldsymbol{H}(\beta)\right]^{-1} = \left\{-\left[\frac{\partial^2 L(\beta)}{\partial\beta\cdot\partial\beta^T}\right]\right\}^{-1}. \tag{A.2}$$

$\boldsymbol{V}^{\text{asym}}$ denotes the asymptotic variance-covariance matrix for the unscaled model, while $\widetilde{\boldsymbol{V}}^{\text{asym}}$ represents the asymptotic variance-covariance matrix for the scaled model in equation (7).

It is straightforward to show that:

$$\begin{aligned}
\boldsymbol{V}^{\text{asym}} &= \frac{1}{\det(\boldsymbol{I}^F)}\begin{bmatrix} \frac{\partial^2 L}{\partial\beta_1\beta_1^T} & -\frac{\partial^2 L}{\partial\beta_0\partial\beta_1} \\ -\frac{\partial^2 L}{\partial\beta_1\partial\beta_0} & \frac{\partial^2 L}{\partial\beta_0\beta_0^T}\end{bmatrix} \\
&= \frac{1}{\det(\boldsymbol{I}^F)}\begin{bmatrix} \sum_{i=1}^n(\log(x_i))^2\exp(\beta_0+\beta_1\log(x_i)) & -\sum_{i=1}^n\log(x_i)\exp(\beta_0+\beta_1\log(x_i)) \\ -\sum_{i=1}^n\log(x_i)\exp(\beta_0+\beta_1\log(x_i)) & \sum_{i=1}^n\exp(\beta_0+\beta_1\log(x_i))\end{bmatrix} \\
&= \frac{\begin{bmatrix} \sum_{i=1}^n(\log(x_i))^2\exp(\beta_0+\beta_1\log(x_i)) & -\sum_{i=1}^n\log(x_i)\exp(\beta_0+\beta_1\log(x_i)) \\ -\sum_{i=1}^n\log(x_i)\exp(\beta_0+\beta_1\log(x_i)) & \sum_{i=1}^n\exp(\beta_0+\beta_1\log(x_i))\end{bmatrix}}{\frac{\partial^2 L}{\partial\beta_0\beta_0^T}\frac{\partial^2 L}{\partial\beta_1\beta_1^T} - \left(\frac{\partial^2 L}{\partial\beta_0\partial\beta_1}\right)^2}.
\end{aligned} \tag{A.3}$$

Following results in Section 3.1, we have $\beta_0 = \tilde{\beta}_0 - \log(S)$ and $\beta_1 = \tilde{\beta}_1$. This allows us to show that $\det(\widetilde{\boldsymbol{I}^F}) = S^2\det(\boldsymbol{I}^F)$:

$$
\begin{aligned}
\det(\widetilde{\boldsymbol{I}^F}) &= \frac{\partial^2 \widetilde{L}}{\partial \widetilde{\beta}_0^2}\frac{\partial^2 \widetilde{L}}{\partial \widetilde{\beta}_1^2} - \left(\frac{\partial^2 \widetilde{L}}{\partial \widetilde{\beta}_0 \partial \widetilde{\beta}_1}\right)^2 \\
&= \sum_{i=1}^n \exp(\widetilde{\beta}_0 + \widetilde{\beta}_1 \log(x_i)) \cdot \sum_{i=1}^n (\log(x_i))^2 \exp(\widetilde{\beta}_0 + \widetilde{\beta}_1 \log(x_i)) \\
&- \sum_{i=1}^n \log(x_i) \exp(\widetilde{\beta}_0 + \widetilde{\beta}_1 \log(x_i)) \cdot \sum_{i=1}^n (\log(x_i))^2 \exp(\widetilde{\beta}_0 + \widetilde{\beta}_1 \log(x_i)) \\
&= \sum_{i=1}^n \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \cdot \sum_{i=1}^n (\log(x_i))^2 \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \\
&- \sum_{i=1}^n \log(x_i) \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \cdot \sum_{i=1}^n (\log(x_i))^2 \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \\
&= S^2 \left[\frac{\partial^2 L}{\partial \beta_0 \beta_0^T}\frac{\partial^2 L}{\partial \beta_1 \beta_1^T} - \left(\frac{\partial^2 L}{\partial \beta_0 \partial \beta_1}\right)^2\right]. \\
&= S^2 \det(\boldsymbol{I}^F)
\end{aligned}
\tag{A.4}
$$

We apply this result to the definition of $\widetilde{\boldsymbol{V}}^{\text{asym}}$:

$$
\begin{aligned}
\widetilde{\boldsymbol{V}}^{\text{asym}} &= \frac{1}{\det(\widetilde{\boldsymbol{I}^F})} \begin{bmatrix} \sum_{i=1}^n (\log(x_i))^2 \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) & -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \\ -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) & \sum_{i=1}^n \exp(\beta_0 + \log(S) + \beta_1 \log(x_i)) \end{bmatrix} \\
&= \frac{S \begin{bmatrix} \sum_{i=1}^n (\log(x_i))^2 \exp(\beta_0 + \beta_1 \log(x_i)) & -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \beta_1 \log(x_i)) \\ -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \beta_1 \log(x_i)) & \sum_{i=1}^n \exp(\beta_0 + \beta_1 \log(x_i)) \end{bmatrix}}{S^2 \left[\frac{\partial^2 L}{\partial \beta_0 \beta_0^T}\frac{\partial^2 L}{\partial \beta_1 \beta_1^T} - \left(\frac{\partial^2 L}{\partial \beta_0 \partial \beta_1}\right)^2\right]} \\
&= \frac{\begin{bmatrix} \sum_{i=1}^n (\log(x_i))^2 \exp(\beta_0 + \beta_1 \log(x_i)) & -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \beta_1 \log(x_i)) \\ -\sum_{i=1}^n \log(x_i) \exp(\beta_0 + \beta_1 \log(x_i)) & \sum_{i=1}^n \exp(\beta_0 + \beta_1 \log(x_i)) \end{bmatrix}}{S \left[\frac{\partial^2 L}{\partial \beta_0 \beta_0^T}\frac{\partial^2 L}{\partial \beta_1 \beta_1^T} - \left(\frac{\partial^2 L}{\partial \beta_0 \partial \beta_1}\right)^2\right]}.
\end{aligned}
\tag{A.5}
$$

Equations (A.3) and (A.5) immediately yield that $\boldsymbol{V}^{\text{asym}} = S\widetilde{\boldsymbol{V}}^{\text{asym}}$. ■

Since the variance is scaled by S, in the case of test of a restriction on β_1, the Wald test statistic calculated from the scaled data ($\widetilde{W}$) is inversely proportional to the Wald test statistic calculated from the unscaled data:

$$\widetilde{W} = \frac{\left(\widetilde{\hat{\beta}}_1 - \beta_1^{\text{restricted}}\right)^2}{\widetilde{V}^{\text{asym}}} = \frac{\left(\hat{\beta}_1 - \beta_1^{\text{restricted}}\right)^2}{SV^{\text{asym}}} = \frac{1}{S}W. \tag{A.6}$$

This result extends to tests of multiple restrictions on the vector $\boldsymbol{\beta}$ (Greene, 2012):

$$W^{\text{multiple restrictions}} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{restricted}})^T \widehat{\boldsymbol{COV}}^{\text{asym}\,-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{restricted}}), \tag{A.7}$$

where

$$\widehat{\boldsymbol{COV}}^{\text{asym}} = \begin{bmatrix} \frac{\partial^2 L^{H_0}}{\partial \beta_1^2} & \frac{\partial^2 L^{H_0}}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 L^{H_0}}{\partial \beta_1 \partial \beta_n} \\ \frac{\partial^2 L^{H_0}}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 L^{H_0}}{\partial \beta_2^2} & \cdots & \frac{\partial^2 L^{H_0}}{\partial \beta_2 \partial \beta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L^{H_0}}{\partial \beta_n \partial \beta_1} & \frac{\partial^2 L^{H_0}}{\partial \beta_n \partial \beta_2} & \cdots & \frac{\partial^2 L^{H_0}}{\partial \beta_n^2} \end{bmatrix}. \tag{A.8}$$

By Corollary 1.1 (in appendix B.1), since each element of the Hessian matrix $\boldsymbol{H}(\beta)$ is scaled by S when the dependent variable y_i is scaled by S: $\widetilde{\widehat{COV}}^{\text{asym}} = S \cdot \widehat{COV}^{\text{asym}}$, we have:

$$\widetilde{W}^{\text{multiple restrictions}} = \frac{1}{S} W^{\text{multiple restrictions}}. \tag{A.9}$$

Since the hypothesis tests involving $\widetilde{W}$ and W compare both statistics to the same critical value, the result of the test depends on the scale of S.

A.2 Wald Statistics Calculated from Robust Standard Errors

An alternative approach to calculating the variance-covariance matrix is to use the residuals from an estimation that returns Huber-White robust standard errors.[10] Consider a Huber-

[10] See Huber (1967) and White (1982).

White bias-corrected variance estimator $\boldsymbol{V}^{\text{H-W}}$ defined as follows:

$$\boldsymbol{V}^{\text{H-W}} = \left[\boldsymbol{I}^F(\beta)\right]^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X} \left[\boldsymbol{I}^F(\beta)\right]^{-1} = \left[\boldsymbol{I}^F(\beta)\right]^{-1} \boldsymbol{X}^T \operatorname{diag}(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon}_n^2) \boldsymbol{X} \left[\boldsymbol{I}^F(\beta)\right]^{-1}, \tag{A.10}$$

where $\boldsymbol{\Sigma} = \boldsymbol{X}^T \operatorname{diag}(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon} n^2) \boldsymbol{X}$ is the *correction matrix* based on raw residuals (Maas and Hox, 2004). The raw residuals vector, $\boldsymbol{\epsilon} = (y_1 - \exp(x_1'\beta), \ldots, y_n - \exp(x_n'\beta))^\top \in \mathbb{R}^n$, consists of the differences between observed values y_i and their corresponding predicted values $\exp(x_i'\beta)$. Let $\boldsymbol{x}_i \in \mathbb{R}^p$ for $i = 1, ..., n$ represent the explanatory variable vectors, and let $\mathbf{1}_n$ be an $n \times 1$ vector of ones. The model matrix $\boldsymbol{X} \in \mathbb{R}^{n \times (p+1)}$ is defined as the concatenation of $\mathbf{1}_n$ and the matrix formed by taking each $\boldsymbol{x}_i$ as a row. The diagonal matrix $\text{diag}_{ij} \in \mathbb{R}^{n \times n}$ has the squared residuals on its main diagonal and zeros on the off-diagonal elements. The $(p+1) \times (p+1)$ matrix $\boldsymbol{\Sigma}$ accounts for the influence of the raw residuals on the standard errors of the estimated parameters in the space $\Upsilon \subseteq \mathbb{R}^p$.[11] The *degrees-of-freedom* adjustment is unaffected by the variable scale:

$$\boldsymbol{V}^{\text{H-W, adjusted}} = \frac{n}{n-1} \boldsymbol{V}^{\text{H-W}}, \tag{A.11}$$

where n is the number of observations in the model. Note that since the y-variables are scaled by $S \neq 1$, and $\log(S)$ is absorbed into the constant (thereby incorporating S into the exponential function), we have $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{y}_1 - \exp(x_1'\widetilde{\beta}), \ldots, \widetilde{y}_n - \exp(x_n'\widetilde{\beta}))^\top = (S \cdot y_1 - S \cdot \exp(x_1'\beta), \ldots, S \cdot y_n - S \cdot \exp(x_n'\beta))^\top = S \cdot \boldsymbol{\epsilon}$.

Therefore, given Equations (A.3), (A.5), (A.10) and (A.11), we have:

[11] For the model specified in equation (2), which follows the standard PPML-gravity framework, it is easy to show that the parameter vector $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_p\}^T$ remains invariant under both x-variable and y-variable scaling across the entire parameter space $\Upsilon \subseteq \mathbb{R}^p$.

$$
\begin{aligned}
\boldsymbol{V}^{\text{H-W, adjusted}} &= \frac{n}{n-1}\left\{\boldsymbol{V}^{\text{asym}}\boldsymbol{X}^T \operatorname{diag}(\hat{\epsilon}_1^2,\ldots,\hat{\epsilon}_n^2)\boldsymbol{X}\boldsymbol{V}^{\text{asym}}\right\} \\
&= \frac{n}{n-1}\left\{\frac{1}{S}\left[\boldsymbol{I}(\beta)\right]^{-1}\boldsymbol{X}^T \operatorname{diag}(S^2\hat{\epsilon}_1^2,\ldots,\hat{S^2\epsilon}_n^2)\boldsymbol{X}\frac{1}{S}\left[\boldsymbol{I}(\beta)\right]^{-1}\right\} \\
&= \frac{n}{n-1}\left\{\left[\widetilde{\boldsymbol{I}}(\widetilde{\beta})\right]^{-1}\boldsymbol{X}^T \operatorname{diag}(S^2\hat{\epsilon}_1^2,\ldots,S^2\hat{\epsilon}_n^2)\boldsymbol{X}\left[\widetilde{\boldsymbol{I}}(\widetilde{\beta})\right]^{-1}\right\} \\
&= \frac{n}{n-1}\left\{\widetilde{\boldsymbol{V}}^{\text{asym}}\boldsymbol{X}^T \operatorname{diag}(\widetilde{\hat{\epsilon}}_1^2,\ldots,\widetilde{\hat{\epsilon}}_n^2)\boldsymbol{X}\widetilde{\boldsymbol{V}}^{\text{asym}}\right\} \\
&= \widetilde{\boldsymbol{V}}^{\text{H-W, adjusted}}.
\end{aligned}
\tag{A.12}
$$

That is, the robust variance estimator does not change with $S \neq 1$ on dependent variables, and neither do the Wald statistics, as can be inferred directly using equation (A.6).[12] This invariance is an algebraic property of the sandwich construction and does not require heteroskedasticity to be present. Under the transformation $y_i \mapsto Sy_i$, the information matrix is multiplied by S, so each inverse-information, or "bread," term is divided by S. At the same time, the residual-based middle matrix, or "meat," is multiplied by S^2 because each residual is multiplied by S. These effects cancel:

$$\left(\frac{1}{S}\right) S^2 \left(\frac{1}{S}\right) = 1.$$

It follows that the Huber–White variance-covariance matrix does not change with the scale of the dependent variable, and neither does the corresponding Wald statistic.

The empirical application in Section 6 confirms the same result. As shown in Table 2, the model-based Wald statistic changes in direct proportion to the scale of export values, while the robust Wald statistic is unchanged.

[12]Note that the k x 1 vector x_t (where $t = \{1, ..., n\}$) of fixed variables are strictly exogenous, while $[x_1, x_2, ..., x_n]^T$ remains unaffected by the scaling choices.

B Extension to Multivariate Settings

In this section we show that the scaling results for the Hessian and the associated variance matrices extend to multivariate settings with multiple parameters. These results are structured as Corollaries to Theorem 1. They imply the corresponding multivariate extension for Wald-type inference and establish the scaling properties of the information matrices that support the extension of LM and related test statistics.

B.1 Corollary 1.1 and Proof

Corollary 1.1 *Let $L(\beta)$ be the twice-differentiable log-likelihood function:*

$$\begin{aligned} L(\beta) = &- \sum_{i}^{n} \exp[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i) + ... + \beta_m \log(w_i)] \\ &+ \sum_{i}^{n} y_i[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i) + ... + \beta_m \log(w_i)], \end{aligned} \tag{B.1}$$

where $\beta_0, \beta_1, ..., \beta_m$ are parameters, and $x_i, z_i, ..., w_i$ are observed from the data. Consider a scalar $S \neq 1$ on y_i, which yields the following log-likelihood function:

$$\begin{aligned} \widetilde{L}(\tilde{\beta}) = &- \sum_{i}^{n} \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i) + \tilde{\beta}_2 \log(z_i) + ... + \tilde{\beta_m} \log(w_i)] \\ &+ \sum_{i}^{n} (Sy_i)[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i) + \tilde{\beta}_2 \log(z_i) + ... + \tilde{\beta_m} \log(w_i)], \end{aligned} \tag{B.2}$$

where $\widetilde{\mathcal{B}} \equiv \{\tilde{\beta}_0, \tilde{\beta}_1, ..., \tilde{\beta_m}\}$ are the parameters given the scale, and the relationship between the original parameters $\mathcal{B}$ and the scaled parameters $\widetilde{\mathcal{B}}$ is as follows:

$$\begin{cases} \beta_0 = \widetilde{\beta}_0 - \log(S) \\ \beta_1 = \widetilde{\beta}_1 \\ \beta_2 = \widetilde{\beta}_2 \\ \vdots \\ \beta_m = \widetilde{\beta}_m, \end{cases} \tag{B.3}$$

then the Hessian matrix of log-likelihood function $\widetilde{L}(\widetilde{\beta})$*, denoted by* $\boldsymbol{H}(\widetilde{\beta})$*, is related to the original Hessian matrix* $\boldsymbol{H}(\beta)$ *by:*

$$\widetilde{\boldsymbol{H}}(\tilde{\beta}) = S \cdot \boldsymbol{H}(\beta). \tag{B.4}$$

Corollary 1.1 asserts that each element of the Hessian matrix $\boldsymbol{H}(\beta)$ (in the PPML function) is scaled by S when the dependent variable y_i is scaled by S.

Proof. Using the chain rule, it is easy to show that the Hessian matrix of the log-likelihood function, for this more general case, is:

$$\boldsymbol{H}(\beta) = \frac{\partial^2 L^{H_0}(\beta)}{\partial \beta \cdot \partial \beta^T} = -\sum_{i=1}^{n} \exp(\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i) + ... + \beta_m \log(w_i)) \cdot \boldsymbol{M}_i, \tag{B.5}$$

where $\boldsymbol{M}_i$ (corresponding to each observation i in the data) is a matrix of the second-order partial derivatives with respect to model parameters $\mathcal{B} \equiv \{\beta_0, \beta_1, ..., \beta_m\}$ as follows:

$$\boldsymbol{M}_i = \begin{bmatrix} 1 & \log(x_i) & \log(z_i) & \cdots & \log(w_i) \\ \log(x_i) & (\log(x_i))^2 & (\log(x_i)\log(z_i)) & \cdots & (\log(x_i)\log(w_i)) \\ \log(z_i) & (\log(z_i)\log(x_i)) & (\log(z_i))^2 & \cdots & (\log(z_i)\log(w_i)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \log(w_i) & (\log(w_i)\log(x_i)) & (\log(w_i)\log(z_i)) & \cdots & (\log(w_i))^2 \end{bmatrix}. \tag{B.6}$$

Note that the scaling factor $S \neq 1$ on y_i does not affect $\boldsymbol{M}_i$, but only the first term of the right-hand side in equation (B.5), where the relationships between the original parameters $\{\beta_0, \beta_1, \ldots, \beta_m\}$ and the scaled parameters $\{\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_m\}$ are given by $\beta_0 = \tilde{\beta}_0 - \log(S)$ and $\beta_j = \tilde{\beta}_j$ for $j = 1, 2, \ldots, m$. Hence, we have

$$\begin{aligned} \widetilde{\boldsymbol{H}}(\widetilde{\beta}) = \frac{\widetilde{\partial^2 L^{H_0}}(\beta)}{\partial\widetilde{\beta} \cdot \partial\widetilde{\beta}^T} &= -\sum_{i=1}^{n} \exp(\widetilde{\beta}_0 + \widetilde{\beta}_1 \log(x_i) + \widetilde{\beta}_2 \log(z_i) + ... + \widetilde{\beta_m} \log(w_i)) \cdot \boldsymbol{M}_i \\ &= -\sum_{i=1}^{n} \exp(\beta_0 + \log(S) + \beta_1 \log(x_i) + \beta_2 \log(z_i) + ... + \beta_m \log(w_i)) \cdot \boldsymbol{M}_i \\ &= -S \sum_{i=1}^{n} \exp(\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i) + ... + \beta_m \log(w_i)) \cdot \boldsymbol{M}_i \\ &= S \cdot \boldsymbol{H}(\beta). \end{aligned} \tag{B.7}$$

■

B.2 Corollary 1.2 and Proof

Corollary 1.2 *Given the conditions of Corollary 1.1, and assuming the Fisher Information Matrix $\boldsymbol{I}^F$, defined over a commutative ring with identity R, is invertible (nonsingular), we can generalize the result in equation (A.12) to accommodate multiple independent variables $\boldsymbol{x}_i \in \mathbb{R}^n$, where $i \geq 2$. This applies to a parameter set $\boldsymbol{\mathcal{B}} = \beta_0, \beta_1, ..., \beta_n{}^T$ within the parameter space $\Gamma \subseteq \mathbb{R}^n$.*

Corollary 1.2 asserts that the results are applicable for joint restrictions on multiple variables.

Proof. We start with the 'large' model introduced in equation (9):

$$L^{H_a}(\beta) = -\sum_{i}^{n} \exp[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log z_i] + \sum_{i}^{n} y_i[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log z_i], \tag{B.8}$$

where $z_i \subseteq \boldsymbol{x}_i \in \mathbb{R}^n$ is the additional independent variable, β_2 is its corresponding model parameter.[13]

Since equation (2) is twice-differentiable, the Hessian matrix is given by:

$$\boldsymbol{H}(\beta) = \begin{bmatrix} \frac{\partial^2 L^{H_a}}{\partial \beta_0 \beta_0^T} & \frac{\partial^2 L^{H_a}}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L^{H_a}}{\partial \beta_0 \partial \beta_2} \\ \frac{\partial^2 L^{H_a}}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 L^{H_a}}{\partial \beta_1 \beta_1^T} & \frac{\partial^2 L^{H_a}}{\partial \beta_1 \partial \beta_2} \\ \frac{\partial^2 L^{H_a}}{\partial \beta_2 \partial \beta_0} & \frac{\partial^2 L^{H_a}}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 L^{H_a}}{\partial \beta_2 \beta_2^T} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \tag{B.9}$$

where each entry is a second order partial derivative of the log-likelihood function with respect to the parameters β_0, β_1, and β_2. This is a symmetric matrix because the order of differentiation does not matter.

Given invertibility, the inverse of the negative Hessian can be simplified as:

$$\boldsymbol{V} = \left(\boldsymbol{I}^F\right)^{-1} = [-\boldsymbol{H}(\beta)]^{-1} = \left\{ - \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \right\}^{-1} = -\frac{1}{\det(\boldsymbol{I}^F)} \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^T, \tag{B.10}$$

where C_{ij} is the cofactor of a_{ij} in the matrix $\boldsymbol{H}(\beta)$, $\det(\boldsymbol{I}^F)$ is the determinant of the Fisher information matrix. In this case, we apply the Rule of Sarrus for the calculation of the determinant of variance-covariance matrix $\boldsymbol{V}$:

[13] We eliminate prime notation used originally in the large model because they are unnecessary here.

$$\begin{aligned}\det(\boldsymbol{I}^F) &= a_{11}\cdot C_{11} - a_{12}\cdot C_{12} + a_{13}\cdot C_{13} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).\end{aligned} \tag{B.11}$$

From Corollary 1.1, it can be shown that each element a_{ij} is scaled by $S \neq 1$, hence:

$$\begin{aligned}\det(\widetilde{\boldsymbol{I}^F}) &= \widetilde{a}_{11}(\widetilde{a}_{22}\widetilde{a}_{33} - \widetilde{a}_{23}\widetilde{a}_{32}) - \widetilde{a}_{12}(\widetilde{a}_{21}\widetilde{a}_{33} - \widetilde{a}_{23}\widetilde{a}_{31}) + \widetilde{a}_{13}(\widetilde{a}_{21}\widetilde{a}_{32} - \widetilde{a}_{22}\widetilde{a}_{31}) \\ &= S^3\cdot[a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})] \\ &= S^3\cdot\det(\boldsymbol{I}^F).\end{aligned} \tag{B.12}$$

By the Cayley–Hamilton theorem shown in Cayley (1858), we have

$$A^n - \mathrm{tr}_1 A^{n-1} + \ldots + (-1)^{n-1}\mathrm{tr}_{n-1}A + (-1)^n\det(A)I_{\mathrm{nxn}} = \mathbf{0}_{\mathrm{nxn}}, \tag{B.13}$$

where $A = \boldsymbol{I}^F = \{a_{ij}\}_{\mathrm{nxn}}$ is a n x n square matrix, satisfying its characteristic polynomial; the trace at the k^{th} level, denoted tr_k, is the sum of all principal minors of A of order k. Since A is invertible, we substitute with $n = 3$, and then multiply through by the inverse of A in equation (B.13):

$$A^2 - \mathrm{tr}(A)A + \mathrm{tr}(A^2)\boldsymbol{I}_{\mathrm{nxn}} - \det(A)A^{-1} = \mathbf{0}_{\mathrm{nxn}}. \tag{B.14}$$

Then the non-robust asymptotic variance-covariance matrix $\boldsymbol{V}$ can be written as follows:

$$\boldsymbol{V}^{\mathrm{asym}} = \left(\boldsymbol{I}^F\right)^{-1} = \frac{1}{\det(\boldsymbol{I}^F)}\left\{\boldsymbol{I}^{F^2} - \mathrm{tr}(\boldsymbol{I}^F)\boldsymbol{I}^F + \mathrm{tr}(\boldsymbol{I}^{F^2})\boldsymbol{I}_{\mathrm{nxn}}\right\}, \tag{B.15}$$

where $\mathrm{tr}(\boldsymbol{I}^F)$ is the trace of the matrix $\boldsymbol{I}^F$.

Since each element of the variance-covariance matrix is scaled by a factor $S \neq 1$, the effect on both the squares of the matrices $\boldsymbol{I}^F$ and the sum of the elements on the main diagonal

will be a scaling by S^2 ($\text{tr}(\widetilde{\boldsymbol{I}^{\boldsymbol{F}}})$ is a linear combination of S and unscaled diagonal elements). Given equation (B.12), we have $\boldsymbol{V}^{\text{asym}} = S\widetilde{\boldsymbol{V}}^{\text{asym}}$. As in equation (A.12), the scaling factor $1/S^2$, which arises from the product $1/S = \widetilde{\boldsymbol{V}}^{\text{asym}}/\boldsymbol{V}^{\text{asym}}$, is canceled out by the S^2 term originating from the raw residuals $\text{diag}(S^2\hat{\epsilon}_1^2, \ldots, S^2\hat{\epsilon}_n^2)$. Thus, $\boldsymbol{V}^{\text{H-W, adjusted}} = \widetilde{\boldsymbol{V}}^{\text{H-W, adjusted}}$.

For n x n Fisher information matrix:

Using the results in Corollary 1.1, we define $\widetilde{\boldsymbol{I}^F}$ ($S \neq 1$) to be the $n \times n$ matrix obtained by multiplying every element in the Fisher information matrix by S:

$$\widetilde{\boldsymbol{I}^F} = \begin{bmatrix} Sv_{11} & Sv_{12} & \ldots & Sv_{1n} \\ Sv_{21} & Sv_{22} & \ldots & Sv_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Sv_{n1} & Sv_{n2} & \ldots & Sv_{nn} \end{bmatrix}. \tag{B.16}$$

Applying Laplace's expansion along the first row:

$$\det(\boldsymbol{I}^F) = \sum_{j=1}^{n} (-1)^{1+j} v_{1j} \boldsymbol{I}_{1j}^F, \tag{B.17}$$

where $\boldsymbol{I}_{1j}^F$ is the determinant of the $(n-1) \times (n-1)$ submatrix that results from deleting the first row and jth column from $\boldsymbol{I}^F$. The determinant of the scaled submatrix is given by:

$$
\begin{aligned}
\det(\widetilde{\boldsymbol{I}^F}) &= \sum_{j=1}^{n}(-1)^{1+j}(\widetilde{v}_{1j})\det(\widetilde{\boldsymbol{I}^F}_{1j}) \\
&= \sum_{j=1}^{n}(-1)^{1+j}(Sv_{1j})\det(\widetilde{\boldsymbol{I}^F}_{1j}) \\
&= \sum_{j=1}^{n}(-1)^{1+j}(Sv_{1j})S^{n-1}\det(\boldsymbol{I}^F_{1j}) \\
&= S^n\sum_{j=1}^{n}(-1)^{1+j}v_{1j}\det(\boldsymbol{I}^F_{1j}) \\
&= S^n\det(\boldsymbol{I}^F).
\end{aligned}
\tag{B.18}
$$

Here we used the fact that $\widetilde{\boldsymbol{I}^F}_{1j}$ is the $(n-1)\times(n-1)$ matrix $\boldsymbol{I}^F_{1j}$ with all of its entries multiplied by S (by applying Corollary 1.1), so the determinant of $\widetilde{\boldsymbol{I}^F}_{1j}$ is $S^{n-1}\det(\boldsymbol{I}^F_{1j})$.

The scaled variance-covariance can be written as:

$$
\widetilde{\boldsymbol{V}} = \begin{bmatrix} Sv_{11} & Sv_{12} & \dots & Sv_{1n} \\ Sv_{21} & Sv_{22} & \dots & Sv_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Sv_{n1} & Sv_{n2} & \dots & Sv_{nn} \end{bmatrix}^{-1} = -\frac{1}{\det(\widetilde{\boldsymbol{I}^F})}\begin{bmatrix} \widetilde{C}_{11} & \widetilde{C}_{12} & \dots & \widetilde{C}_{1n} \\ \widetilde{C}_{21} & \widetilde{C}_{22} & \dots & \widetilde{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{C}_{n1} & \widetilde{C}_{n2} & \dots & \widetilde{C}_{nn} \end{bmatrix}^{T} = -\frac{\mathrm{adj}(\widetilde{\boldsymbol{I}})}{\det(\widetilde{\boldsymbol{I}^F})}.
\tag{B.19}
$$

Since each determinant of the $(n-1)\times(n-1)$ submatrix of $\boldsymbol{I}^F$ ($\det(\boldsymbol{I}^F_{ij})$) is scaled by S^{n-1}, thus each cofactor C_{ij} is scaled by S^{n-1}: $\widetilde{C}_{ij} = (-1)^{i+j}\det(\widetilde{\boldsymbol{I}^F}_{ij}) = S^{n-1}C_{ij}$. It follows that the adjugate is scaled by S^{n-1}: $\mathrm{adj}(\widetilde{\boldsymbol{I}}) = S^{n-1}\cdot\mathrm{adj}(\boldsymbol{I})$. Since the adjugate is scaled by S^{n-1} in the numerator, and the determinant is scaled by S^n; hence, we have $\boldsymbol{V}^{\mathrm{asym}} = S\widetilde{\boldsymbol{V}}^{\mathrm{asym}}$. The remainder of the proof proceeds as in equation (A.12). ■

C Analytical Derivations for LM Test

We now consider the implications of scaling for the LM test of the restriction $H_0 : \beta_2 = 0$ in the 'large' model (equation (9)). Let

$$\theta = (\beta_0, \beta_1, \beta_2)', \qquad \hat{\theta}_R = (\hat{\beta}_0, \hat{\beta}_1, 0)'$$

denote the parameter vector of the 'large' model and the restricted estimator obtained from the 'small' model, respectively.

Because $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy the first-order conditions of the restricted model, the first two score components of the 'large' model, evaluated at $\hat{\theta}_R$, vanish. The only nonzero component is therefore the score associated with the excluded parameter β_2:

$$U(\hat{\theta}_R) = \frac{\partial L^{H_a}(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_R} = \begin{bmatrix} 0 \\ 0 \\ \sum_{i=1}^{n} \log(z_i)\,[y_i - \hat{\mu}_i] \end{bmatrix}, \qquad \hat{\mu}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log x_i),$$

where $\hat{\mu}_i$ denotes the fitted mean under the restricted model. The Hessian of the 'large' model, evaluated at the restricted estimates, is

$$H(\hat{\theta}_R) = -\sum_{i=1}^{n} \hat{\mu}_i \begin{bmatrix} 1 & \log(x_i) & \log(z_i) \\ \log(x_i) & (\log(x_i))^2 & \log(x_i)\log(z_i) \\ \log(z_i) & \log(x_i)\log(z_i) & (\log(z_i))^2 \end{bmatrix}. \tag{C.1}$$

The LM statistic is therefore

$$\text{LM} = U(\hat{\theta}_R)^T [-H(\hat{\theta}_R)]^{-1} U(\hat{\theta}_R). \tag{C.2}$$

Under the rescaling $\tilde{y}_i = S \cdot y_i$ with $S > 0$, the restricted fitted mean in the scaled model satisfies $\tilde{\mu}_i = S \cdot \hat{\mu}_i$. Hence

$$\widetilde{U}(\hat{\theta}_R) = S \cdot U(\hat{\theta}_R), \qquad -\widetilde{H}(\hat{\theta}_R) = S[-H(\hat{\theta}_R)],$$

so that

$$\begin{aligned}
\widetilde{\mathrm{LM}} &= \widetilde{U}(\hat{\theta}_R)^T[-\widetilde{H}(\hat{\theta}_R)]^{-1}\widetilde{U}(\hat{\theta}_R) \\
&= (S \cdot U(\hat{\theta}_R))^T \left[S \cdot (-H(\hat{\theta}_R))\right]^{-1} (S \cdot U(\hat{\theta}_R)) \\
&= S \cdot \mathrm{LM}.
\end{aligned} \tag{C.3}$$

As with the other cases, hypothesis tests based on the conventional LM statistic are sensitive to scaling because the statistic changes with the scale of the dependent variable, whereas the reference critical value does not.

C.1 Heteroskedasticity-Robust LM Statistic

A correct adjustment is to replace the model-based information matrix with the empirical covariance matrix of the score contributions. This yields a heteroskedasticity-robust statistic, because inference is based on the outer product of the score rather than the model-implied information matrix, and therefore does not rely on the Poisson variance assumption $\mathrm{Var}(y_i \mid x_i) = \mu_i(\theta)$. Let

$$w_i = \begin{bmatrix} 1 \\ \log(x_i) \\ \log(z_i) \end{bmatrix}, \qquad s_i(\theta) = w_i \left[y_i - \mu_i(\theta)\right], \qquad \mu_i(\theta) = \exp(\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(z_i)).$$

Define

$$A(\theta) = -\sum_{i=1}^{n} \frac{\partial s_i(\theta)}{\partial \theta'} = \sum_{i=1}^{n} \mu_i(\theta)\, w_i w_i', \qquad B(\theta) = \sum_{i=1}^{n} s_i(\theta) s_i(\theta)'.$$

Partition $\theta = (\eta', \beta_2)'$, where $\eta = (\beta_0, \beta_1)'$, and evaluate all quantities at the restricted estimator $\hat{\theta}_R$. Then the heteroskedasticity-robust LM statistic for testing $H_0 : \beta_2 = 0$ is

$$\mathrm{LM}^{\mathrm{robust}} = U_2(\hat{\theta}_R)^T \left[C\, B(\hat{\theta}_R)\, C^T \right]^{-1} U_2(\hat{\theta}_R), \qquad C = \begin{bmatrix} - A_{2\eta} A_{\eta\eta}^{-1} & 1 \end{bmatrix},$$

where $U_2(\hat{\theta}_R) = \sum_{i=1}^n \log(z_i)\,[y_i - \hat{\mu}_i]$, and

$$A(\hat{\theta}_R) = \sum_{i=1}^n \hat{\mu}_i \begin{bmatrix} 1 & \log(x_i) & \log(z_i) \\ \log(x_i) & (\log(x_i))^2 & \log(x_i)\log(z_i) \\ \log(z_i) & \log(x_i)\log(z_i) & (\log(z_i))^2 \end{bmatrix} = \begin{bmatrix} A_{\eta\eta} & A_{\eta 2} \\ A_{2\eta} & A_{22} \end{bmatrix}, \tag{C.4}$$

where

$$A_{\eta\eta} = \sum_{i=1}^n \hat{\mu}_i \begin{bmatrix} 1 & \log(x_i) \\ \log(x_i) & (\log(x_i))^2 \end{bmatrix}, \quad A_{\eta 2} = \sum_{i=1}^n \hat{\mu}_i \begin{bmatrix} \log(z_i) \\ \log(x_i)\log(z_i) \end{bmatrix},$$

$$A_{2\eta} = \sum_{i=1}^n \hat{\mu}_i \begin{bmatrix} \log(z_i) & \log(x_i)\log(z_i) \end{bmatrix}, \quad A_{22} = \sum_{i=1}^n \hat{\mu}_i (\log(z_i))^2.$$

Under the rescaling $\tilde{y}_i = S \cdot y_i$ with $S > 0$, we have

$$\tilde{s}_i = S \cdot s_i, \qquad \tilde{U}_2 = S \cdot U_2, \qquad \tilde{A} = S \cdot A, \qquad \tilde{B} = S^2 \cdot B,$$

while C is unchanged. Thus

$$\widetilde{\mathrm{LM}}^{\mathrm{robust}} = (S \cdot U_2)^T \left[C(S^2 B) C^T \right]^{-1} (S \cdot U_2) = \mathrm{LM}^{\mathrm{robust}}. \tag{C.5}$$

Hence the heteroskedasticity-robust LM statistic is invariant to multiplicative scaling of the dependent variable.