Making Gravity Great Again

Will Martin

Selected presentation for the International Agricultural Trade Research Consortium's (IATRC's) 2020 Annual Meeting: Economic Implications of COVID-19, December 14-15, 2020, Virtual platform.

Making Gravity Great Again

IATRC Annual Meeting

Will Martin

International Food Policy Research Institute

15 December 2020

Roadmap

- Importance of the Issue
- Econometric problems
- Approaches to estimation
- Monte Carlo estimates of bias

IFPRI

Importance of the question

- Gravity model fits data well– widely used for trade, investment, migration etc
- Many recent suggestions to replace traditional trade policy models
- Lack of agreement on the best estimators
  - Wide differences between results from different estimators

IFPRI

Econometric problems

IFPRI

Econometric challenges

- Limited-Dependent Variable Bias
  - Large shares of trade, investment, migration data are zeros
    - 40% in aggregate trade flows, higher for disaggregate data
      - Creates bias due to non-zero errors
- Combination of nonlinearity & heteroscedasticity
  - Correlation between errors & explanatory variables → bias
    - Inefficient rather than biased with linear models
      - Estimating in logs gives linearity, but then the zeros truncated
- Heteroscedasticity gives different bias with limited-dependent estimators
  - Whether linear or non-linear
- Some zero observations are not zero, just missing

IFPRI

Eaton-Kortum Tobit Model

$$y_i^* = f(x_i, \beta) + u_i$$

Threshold

$$y_i = y_i^* \quad \text{if} \quad y_i^* > \gamma \quad \text{or}$$

$$y_i = 0 \quad \text{if} \quad y_i^* \leq \gamma$$

- Where *y* is a latent variable, *x* an explanatory variable, γ a minimum threshold, and *u* an error term, that differs between observations

IFPRI

Likelihood function for this model

- $L(\beta,\sigma) = \prod_{i=1}^{n_0} \Phi\left(\frac{\gamma - f(x_i,\beta)}{\sigma_i}\right) \prod_{i=n_0+1}^{n} \frac{1}{\sigma_i} \phi\left(\frac{y - f(x_i,\beta)}{\sigma_i}\right)$

- ϕ is the density function for non-limit obsvns

- Φ is the distribution function for limit observations

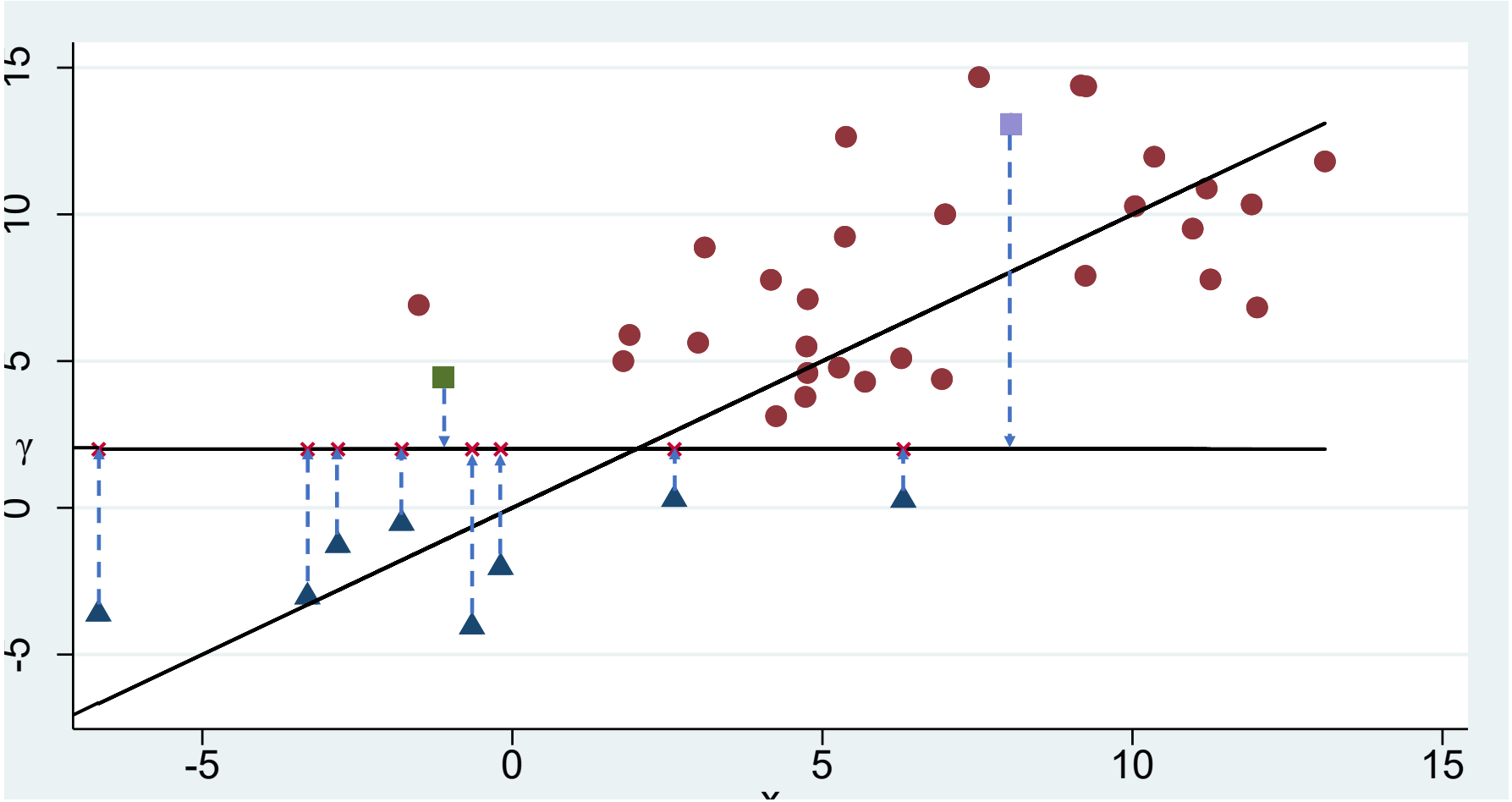
IFPRI

Eaton-Kortum/Carson-Sun Threshold Model

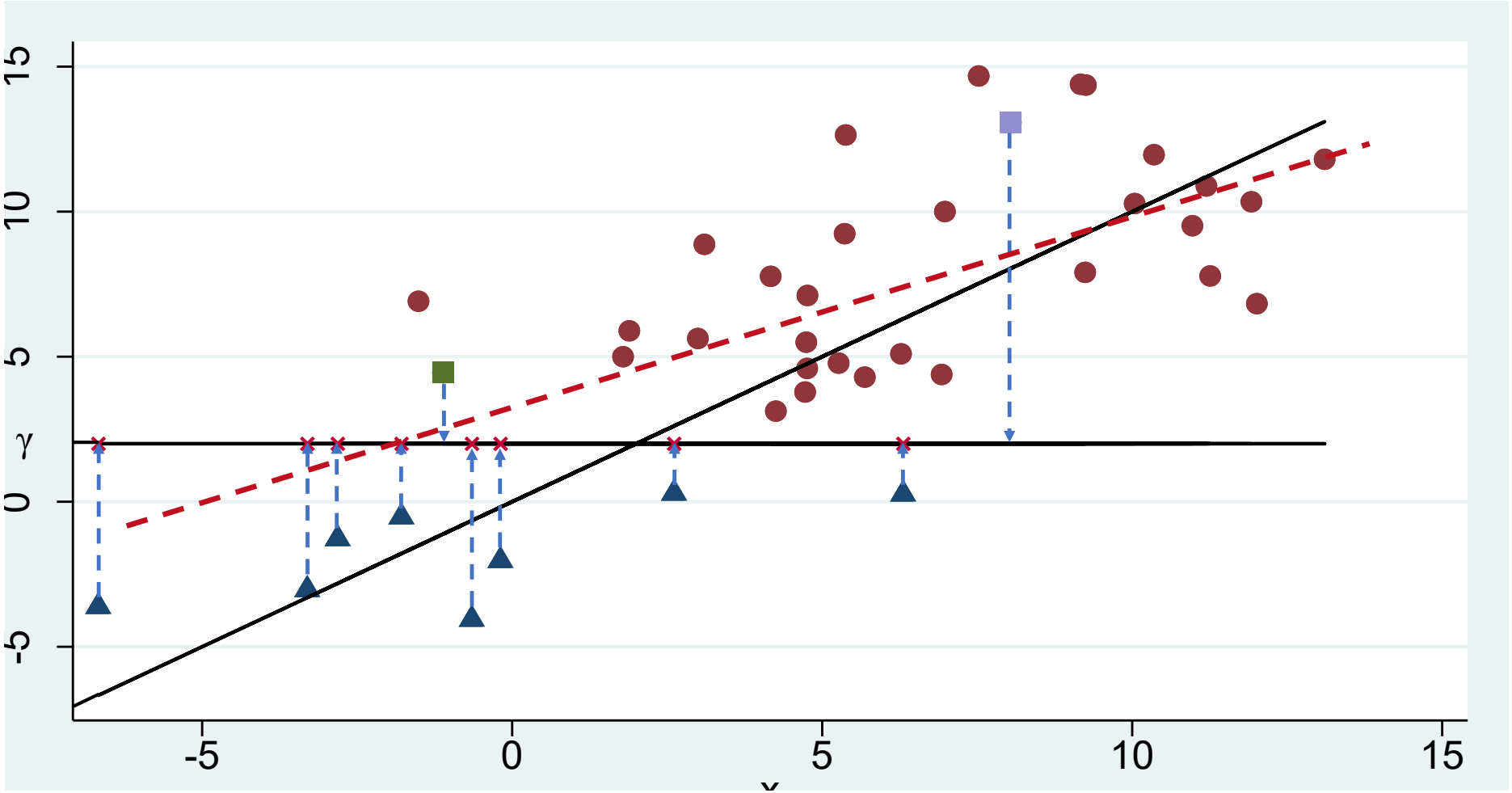
- Threshold often taken to be zero
  - But the lowest profitable value of trade γ >0
- Carson-Sun examine automobiles
  - Smallest new vehicle purchase the cheapest available car
- Eaton-Kortum propose using the lowest observed import to country
  - Carson-Sun show estimator is super-consistent wrt γ
- γ >0 eliminates the zero problem in log models

IFPRI

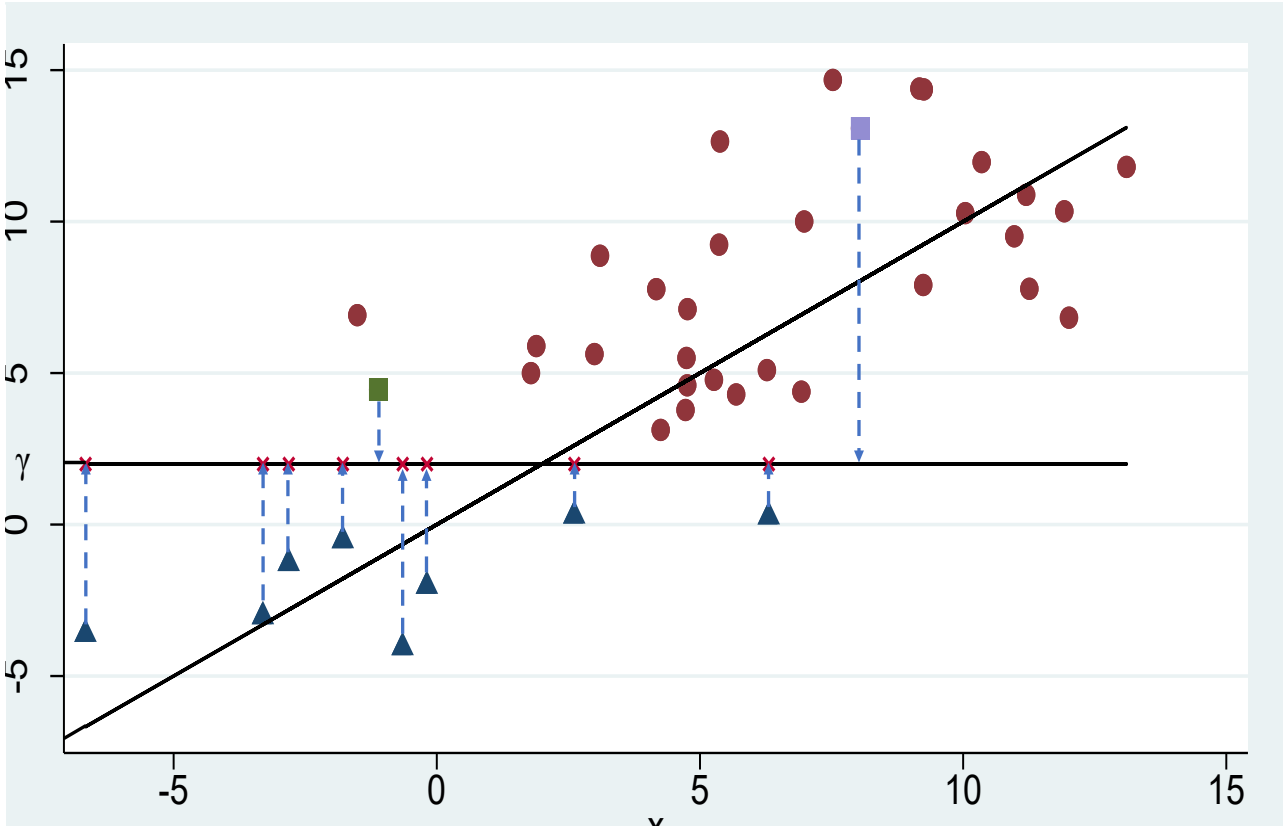
x Values represented by Distn fn, ● s by Density



Limited-Dependent: Downward Bias with OLS



Missing Data Recorded as Zeros



- Biased up if missing data small
- Biased down if missing data large

IFPRI

Estimation

IFPRI

Consider six estimators

- OLS in logs
  - Traditional model– zeros truncated
- PPML
- Generalized Least Squares (for heteroscedasticity)
- Threshold Tobit
- Weighted Tobit
- Eaton-Kortum with heteroscedasticity, EK-H

IFPRI

Dataset

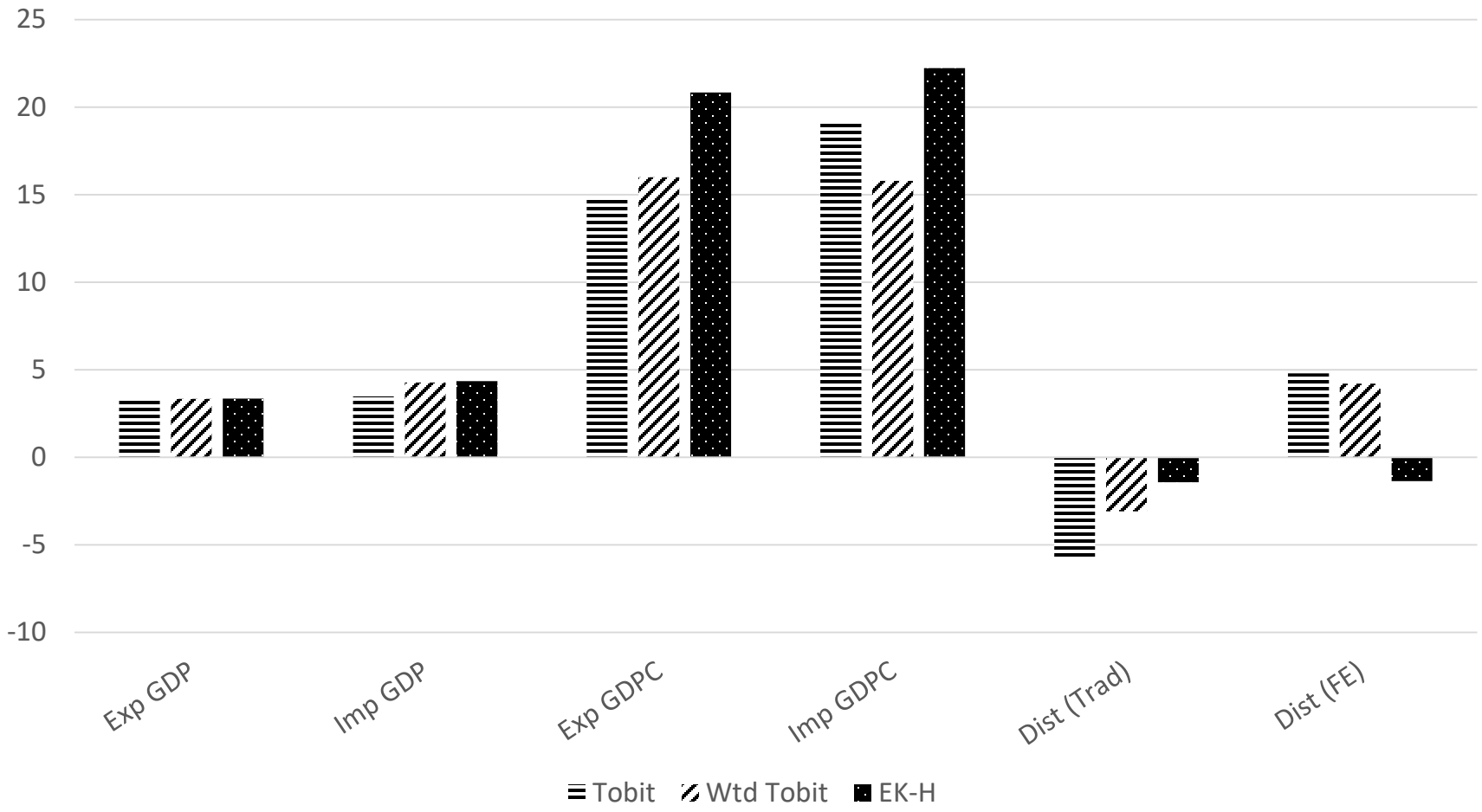
- 136*136 cross-sectional trade matrix by Feenstra, Lipsey & Bowen
  - Used by Santos Silva & Tenreyro & many other studies
    - 48% zeros

- 46 of 136 economies did not report trade to COMTRADE for 1992
  - 2059 of the potential 2070 trade flows between them were zero
    - Clearly no attempt made to impute these data
      - Safer to exclude them

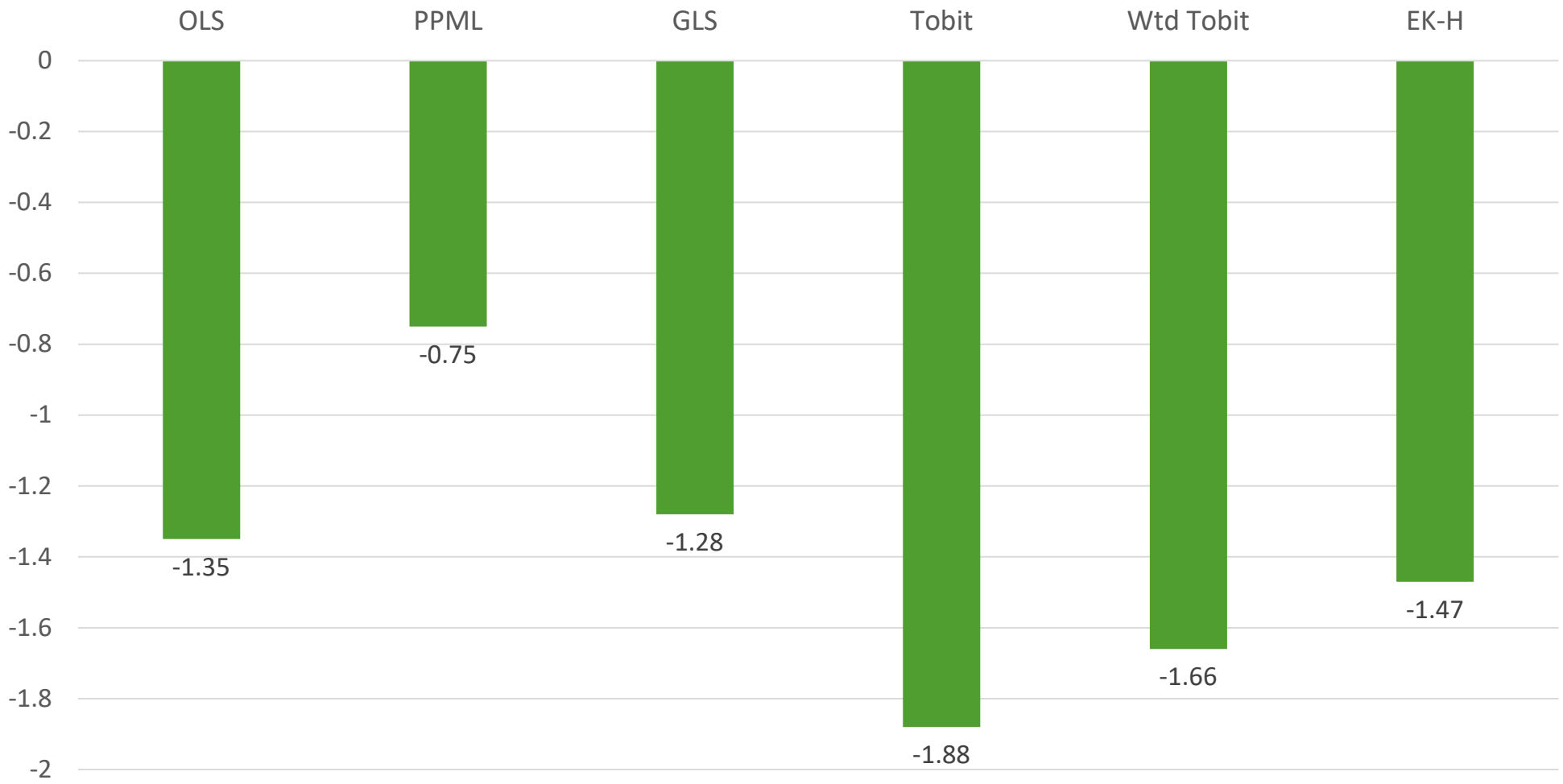
- Resulting 16290 cross-sectional observations with 41% zeros

IFPRI

Effect of adding the missing data as zeros – upward bias



Huge Differences in Estimates: Dist. in FE Gravity



Monte Carlo Analysis

IFPRI

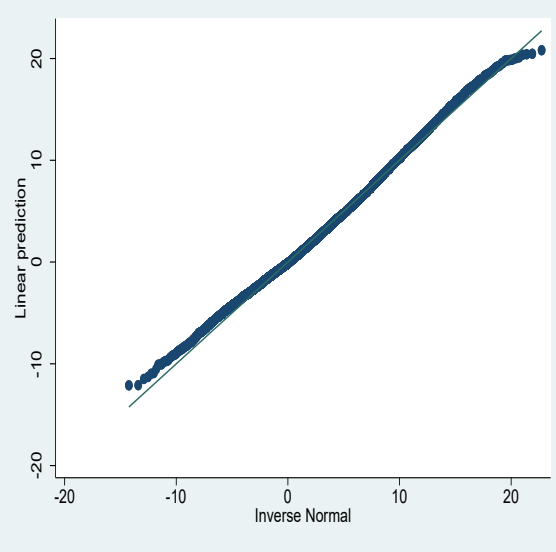
Monte Carlo Analysis misleading unless parameters well-chosen. What matters?

1. The *distribution of predicted* outcomes, including
   (i) Their distribution & whether normal or log-normal,
   (ii) The mean & variance of this distribution
2. *Distribution of residuals* around expected trade outcomes, incl:
   (i) Their distribution,
   (ii) Their standard deviation at any point, σ_i, and
   (iii)The pattern of heteroscedasticity
3. The fraction of *observations at the threshold*

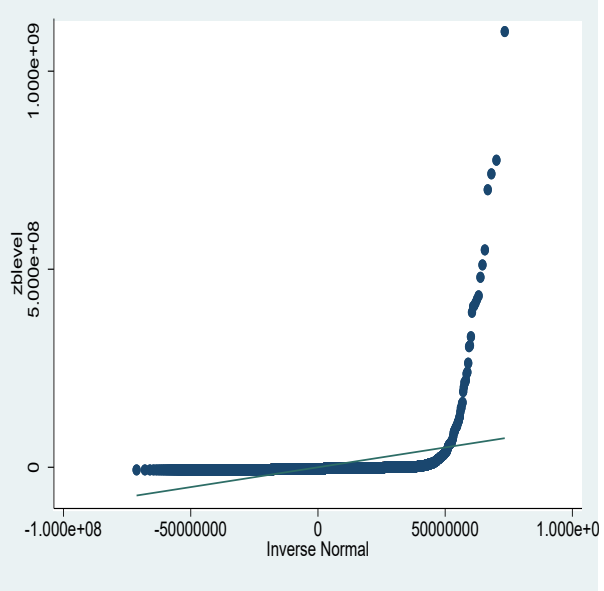
IFPRI

QQ plots 4 Normality

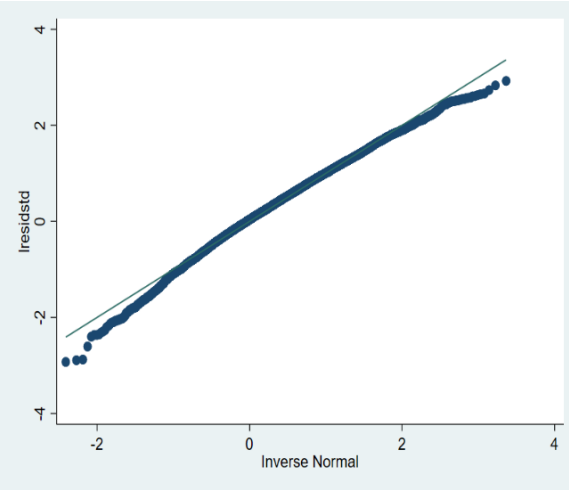
1. Predicted Values, logs



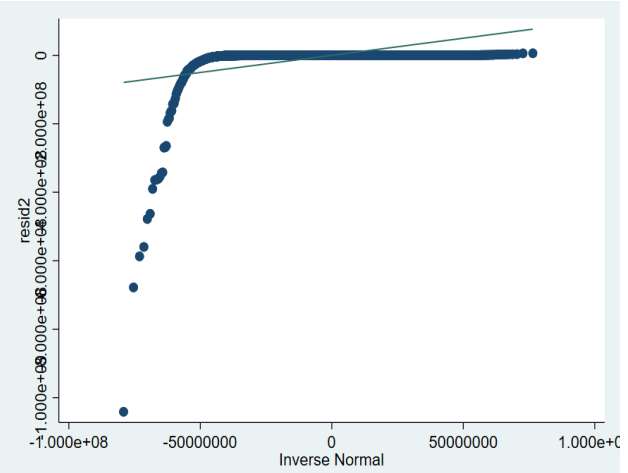
2. Predicted Values, Levels



3. Residuals, logs



4. Residuals, levels



IFPRI

Parameters from 2 candidate models: PPML & EK-H

- *Test model:* $y = 1.x + u$
- x~N(4.2, 4.85) for EK-H Data Gen Process
- x~N(8.2, 2.8) for PPML DGP
- $\sigma_i^2 = e^{(a+by_i)}$ where a= 2.7 & b= –0.115
- Samples of 10,000 repeated 1000 times

IFPRI

Monte Carlo Results

	EK-H Parameters		PPML Parameters	
	$\hat{\beta}$	Std Error	$\hat{\beta}$	Std Error
OLS	0.62	0.01	0.52	0.01
PPML	0.87	0.13	0.78	0.11
Tobit	0.89	0.01	0.90	0.01
Weighted Tobit	0.92	0.03	0.85	0.02
EK-H	0.999	0.01	1.000	0.001

IFPRI

Conclusions

- Need good estimates from the gravity model
  - Potentially serious biases from
    - (i) limited dependent variables
    - (ii) heteroscedasticity & nonlinearity, &
    - (iii) missing values mapped to zeros
- Estimates vary enormously between widely-used estimators
  - Need Monte Carlo analyses to distinguish against known parameters
    - Must use distributions in the relevant range
- EK-H estimator allowing for heteroscedasticity & limited-dependent vbls unbiased
  - Traditional OLS very seriously biased
  - PPML biased around 20%, Tobit & Hetero-Wtd Tobit a little better

IFPRI