A simulation approach for evaluating hedonic wage models' ability to recover marginal values for risk reductions

Xingyi S. Puckett

PhD. Candidate

Center for Environmental and Resource Economic Policy
North Carolina State University

Value of a Statistical Life: Policy Importance

- Total benefit estimates for many federal regulatory programs are driven by point-estimates of the VSL

- Majority of evidence relies on hedonic wage models to estimate the VSL

- Link between VSL estimate and MWTP for risk

Main issues of hedonic wage models

- Measurement error on the job fatal risk
  - Occupation-industry categories are coarsely aggregated
- Omitted variable bias
  - Unobserved worker's characteristics
  - Unobserved job's attributes

- Estimates of MWTP for risk are highly unstable and sensitive to the model specifications

A simulation approach to evaluate the modeling specifications

- Research goal:
  - Using newly improved data to systematically identify the influence of modeling choices on the accuracy of MWTP estimates
    - ➢ Which functional form is appropriate?
    - ➢ Which estimator is more preferable?
    - ➢ Does occupation and/or industry fixed effects help to increase the accuracy of the estimates?

A simulation approach to evaluate the modeling specifications------research overview

- Stage 1:
  - Use a sorting model to simulate a hedonic market equilibrium
  - Provide database to evaluate the hedonic wage models
- Stage 2:
  - compute "true" MWTP to evaluate the performance of reduced form hedonic wage models in the face of omitted variables

Stage 1: Simulating Hedonic Equilibrium ---- hedonic wage equilibrium

- Sorting model
  - Theory of Compensation wage differentials
  - Generalization of Roy sorting model

  - Basic idea:

$$\max \mathrm{EU} = \max\ (1\text{-}P_j)^*\mathrm{U}\ (\overline{y_i} + W_i^j, Z_j\ ;\ \alpha_i) + P_j{}^*\mathrm{U}(\overline{y_i} + \widetilde{W_i^j}).$$

$$(1\text{-}P_j)\mathrm{U}\ (\overline{y_i} + W_i^j, Z_j\ ;\ \alpha_i) = (1\text{-}P_{j+k})\mathrm{U}\ (\overline{y_i} + W_i^j + \delta_i^{j'}, Z_{j'}\ ;\ \alpha_i).$$

$$\widetilde{W_i^{j'}} = W_i^j + \delta_i^{j'} = U^{-1}(Z_j, P_j, W_i^j, Z_{j'}, P_{j'}; \alpha_i)\ \text{-}\overline{y_i}\ .$$

$$W_i^{j'} = \widetilde{W_i^{j'}} + \varepsilon = U^{-1}(Z_j, P_j, W_i^j, Z_{j'}, P_{j'}; \alpha_i)\ \text{-}\overline{y_i} + \varepsilon.$$

$\overline{y}_i$ *: non-labor income*

Z_i *: vector of job attributes (from O*NET data)*

α_i *: vector of preference parameters*

Stage 1: Simulating Hedonic Equilibrium ---- Data (2004-2006)

- Labor Force data:
  - Survey of Income and Program Participation (SIPP)
    - Workers characteristics
- ***New*** Fatal risk data:
  - Census of Fatal Occupation Injuries (CFOI) introduced in 1992 & first utilized in 2001:
    - Complete census of workplace deaths
    - Available by occupation within industry
- ***New*** Job Attributes data :
  - Occupational information Network (O*NET)
    - Detailed description on skill and occupational requirement
    - Available by detailed occupations

Stage 1: Simulating Hedonic Equilibrium -----sorting algorithm

- Model set up (m workers $\xrightarrow{\text{sort into}}$ n jobs):

Workers ----- observation from labor force data

# of workers: ***1,890***

Jobs/firms ----- defined by fatal risk rate, job attributes

# of jobs in 2004: ***650***

# of jobs in 2005: ***657***

# of jobs in 2006: ***671***

hiring capacity: # of workers observed in each job

Worker's Utility: $E(\ln U_i^j) = (1\text{-}P_j)[\ln(W_i^j + \overline{y_i}) + \alpha_i \ln(Z_j)]$

Worker's baseline utility ----- determined by non-labor income

Stage 1: Simulating Hedonic Equilibrium -----sorting algorithm

- Sorting algorithm
  - *Step #1*. Given the job attributes and baseline utility, firms calculate the competitive wage for each worker. Workers are ranked from low wage to high wage. Firm makes offers to fill up the vacancy.

$$\widetilde{U}_i = \mathrm{U}\,(\overline{y_i}) = (1\text{-}P_j)\mathrm{U}\,(\overline{y_i} + \widetilde{W}_{i,t=0}^{j}, Z_j\ ;\ \alpha_i),$$

$$W_{i,t=0}^{j} = U^{-1}(\widetilde{U}_i, Z_j, P_j, ;\ \alpha_i) - \overline{y_i} + \varepsilon$$

  - *Step #2*. Worker updates utility level from best offer and makes rejection.

Stage 1: Simulating Hedonic Equilibrium -----sorting algorithm

– *Step #3*. Firms re-calculate the competitive wages for workers by taking account of worker's current utility level.

- Wage offers not being rejected in previous step remain the same.
- Wage offers being rejected in previous step have to increase.

$$W_{i,t}^{j} = W_{i,t-1}^{j} \text{ or } W_{i,t}^{j} = U^{-1}(\underline{Z_k, P_k}, W_{i,t-1}^{k}, \underline{Z_{j'}, P_{j'}}; \alpha_i) - \overline{y_i} + \varsigma$$

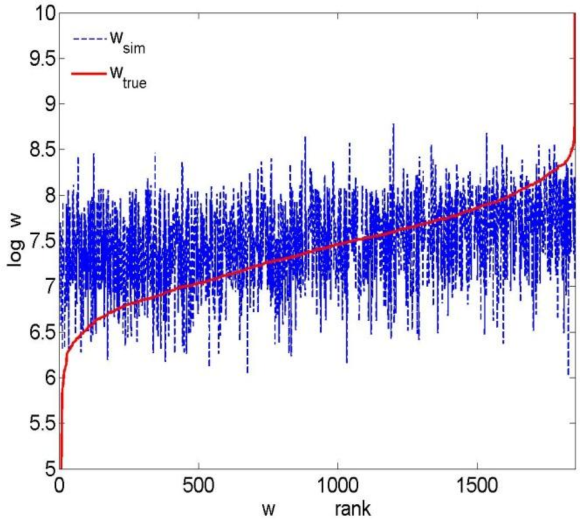
– *Step #4*. The process stops when no rejection is issued in the current round. Market equilibrium occurs.

Stage 1: Simulating Hedonic Equilibrium -----sorting algorithm

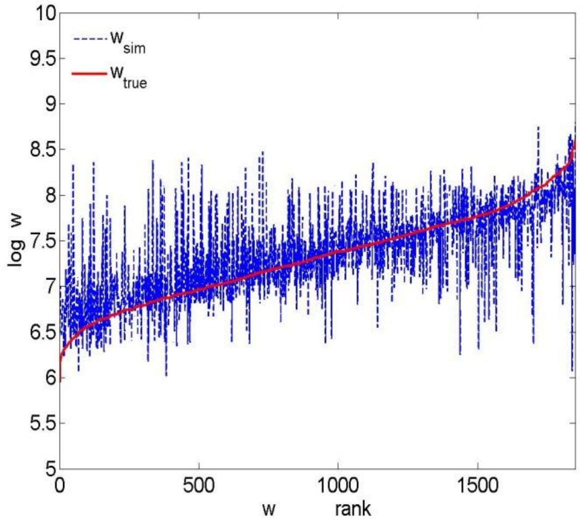
- How is worker's preference determined?
  - Considered as a product of worker's observed and unobserved characteristics such as taste and productivity
  - Calibrating the preference parameters to reproduce the wage pattern in actual labor force data
    - ➢ Capturing wage differentials among workers
  - Starting value for calibration:
    - ➢ 2004: Random Utility model
    - ➢ 2005: optimized results from 2004
    - ➢ 2006: optimized results from 2005

Stage 1: Simulating Hedonic Equilibrium --- simulation results

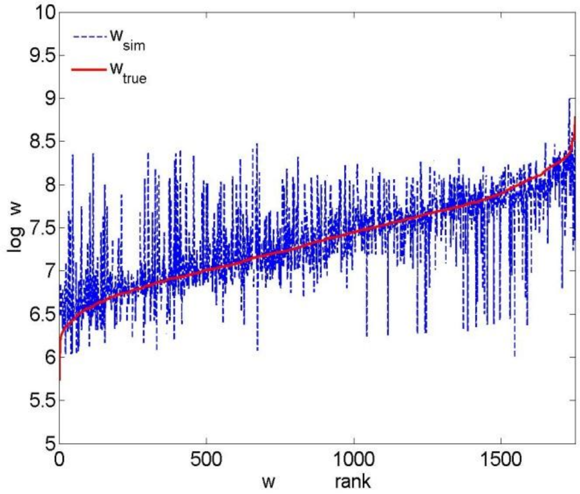
A:



B:



C:



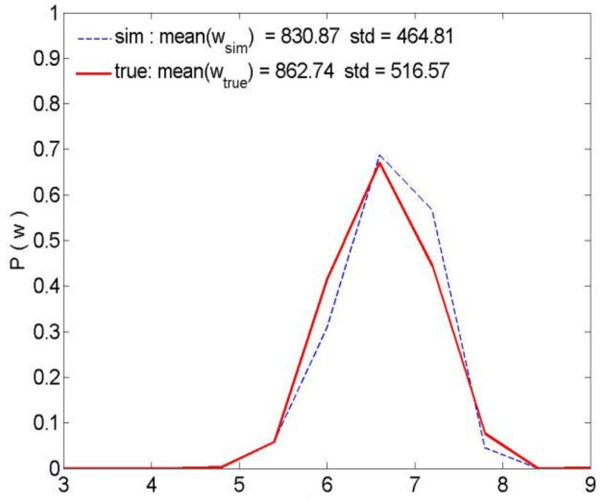
A: Derivative from CDF (year 2004)

B: Derivative from CDF (year 2005)

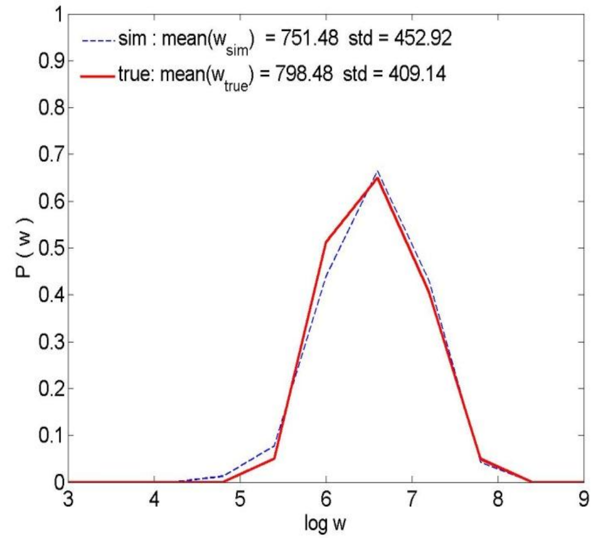
C: Derivative from CDF (year 2006)

Stage 1: Simulating Hedonic Equilibrium --- simulation results

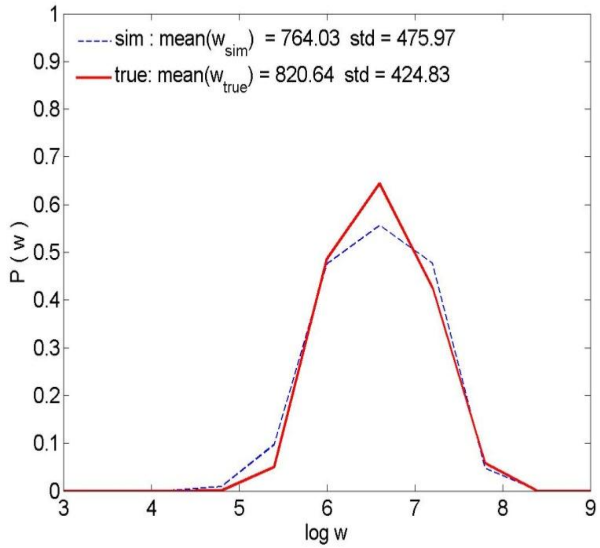
A:



B:



C:



A: Derivative from PDF (year 2004)

B: Derivative from PDF (year 2005)

C: Derivative from PDF (year 2006)

Stage 1: Simulating Hedonic Equilibrium -----simulated database

- Simulated database
  - A set of market equilibrium wage for workers
  - Worker's preference
  - Observation of worker's actual characteristics
  - Observation of worker's job arising from the simulation
  - The job attributes including fatal risk, associated with worker's job

$$\longrightarrow \quad \mathrm{E(U)=max}\ (1\text{-p})\, U_0\, (\bar{y} + \mathrm{W(p)}) + \mathrm{p}\, U_1\, (\bar{y} + \tilde{W}),$$

$$\mathrm{MWTP} \equiv \frac{dw}{dp} = \frac{U_0(\bar{y} + \mathrm{W(p)}) - U_1(\bar{y} + \tilde{W})}{(1-p)U_0'(\bar{y} + \mathrm{W(p)}) + pU_1'(\bar{y} + \tilde{W})},$$

Stage 2: Evaluating hedonic wage model specifications

- Estimate 3,480 individual hedonic wage models

- using combinations of key modeling choices, including:
  - which functional form is used in estimation
  - the approaches to cross-section OLS and panel data estimation
  - the presences of occ. and ind. fixed effect
  - the inclusion of job attributes from O*NET data

Stage 2: Evaluating hedonic wage model specifications

- General Regression Model:

$$f(\text{wage})=\alpha + \beta_1 f(\text{risk}) + \Upsilon_1 \text{injrisk} + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k Occ_k + \sum_{m=1}^{M} \mu_m Ind_m + \varepsilon$$

f(wage) = transformation of wage

f(risk) = transformation fatality risk rate

injrisk = nonfatal injury risk rate

X = set of worker' characteristics

Z = set of jobs' characteristics

Occ = dummy variables indicating occupation of worker

Ind = dummy variables indicating industry of the worker

Model Variations

$$\text{f(wage)} = \alpha + \beta_1 \text{f(risk)} + \Upsilon_1 \text{injrisk} + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k Occ_k + \sum_{m=1}^{M} \mu_m Ind_m + \varepsilon$$

i. 7 estimation functional forms include:

- Linear
- Semi-log
- Double-log
- Linear Box-Cox transformation
- Quadratic Box-Cox transformation
- Linear functional form with quadratic fatal risk term
- Semi-log functional form with quadratic fatal risk term

Model Variations

$$\text{f(wage)}=\alpha + \beta_1 \text{f(risk)} + \Upsilon_1 \text{injrisk} + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k \boxed{Occ_k} + \sum_{m=1}^{M} \mu_m \boxed{Ind_m} + \varepsilon$$

ii. 2 choices for occupation fixed effects:

– K = 0, 6

iii. 2 choices for industry fixed effects:

– L = 0, 9

Model Variations

$$f(\text{wage})=\alpha + \beta_1 f(\text{risk}) + \Upsilon_1 \text{injrisk} + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k Occ_k + \sum_{m=1}^{M} \mu_m Ind_m + \varepsilon$$

iv. 30 choices of O*NET job attributes

- Cognitive skill demand
- Motor skill demand
- Physical skill demand
- Working condition involving hazard exposure
- Social skill

Stage 2: Evaluating hedonic wage model specifications

v. 5 cross-sectional and panel data estimators

- Single OLS (2006)
- Pooled OLS (2004-2006)
- Fixed effect (2004-2006)
- First difference (Δ=1) (2004-2006)
- First difference (Δ=2) (2004-2006)

Model Variations

$$f(\text{wage}) = \alpha + \beta_1 f(\text{risk}) + \Upsilon_1 \text{injrisk} + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k Occ_k + \sum_{m=1}^{M} \mu_m Ind_m + \varepsilon$$

i. 7 estimation functional forms

ii. 2 choices for occupation fixed effects:

- K = 0, 6

iii. 2 choices for industry fixed effects:

- L = 0, 9

iv. 30 choices of O*NET job attributes groups

v. 5 cross-sectional and panel data estimators

=> 3,480 regressions in total

Stage 2: Evaluating hedonic wage model specifications

- Evaluation method
  - difference (bias) for empirical hedonic wage estimate across workers

$$e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i$$

  - computing $\frac{\partial W(P_i)}{\partial P_i}$ for each worker:

$$f(wage) = \alpha + \beta_1 f(risk) + \Upsilon_1 injrisk + \sum_{j=1}^{J} \delta_j X_j + \sum_{k=1}^{K} \varphi_k Z_k + \sum_{l=1}^{L} \mu_k Occ_k + \sum_{m=1}^{M} \mu_m Ind_m + \varepsilon$$

  - computing $MWTP_i$ for each worker:

$$E(U) = \max\ (1-p)\, U_0\, (\bar{y} + W(p)) + p\, U_1\, (\bar{y} + \tilde{W}),$$

$$MWTP \equiv \frac{dw}{dp} = \frac{U_0(\bar{y} + W(p)) - U_1(\bar{y} + \tilde{W})}{(1-p)U_0'(\bar{y} + W(p)) + pU_1'(\bar{y} + \tilde{W})},$$

Stage 2: Evaluating hedonic wage model specifications

- Evaluation method
  - Summary statistics of bias of each hedonic wage model:

$$\beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i} \qquad S_p = \frac{\bar{s}_p}{N^{-1}\sum_i MWTP_i}$$

$\Rightarrow$ # of β_p and S_p : ***3,480***.

An Internal Meta Analysis

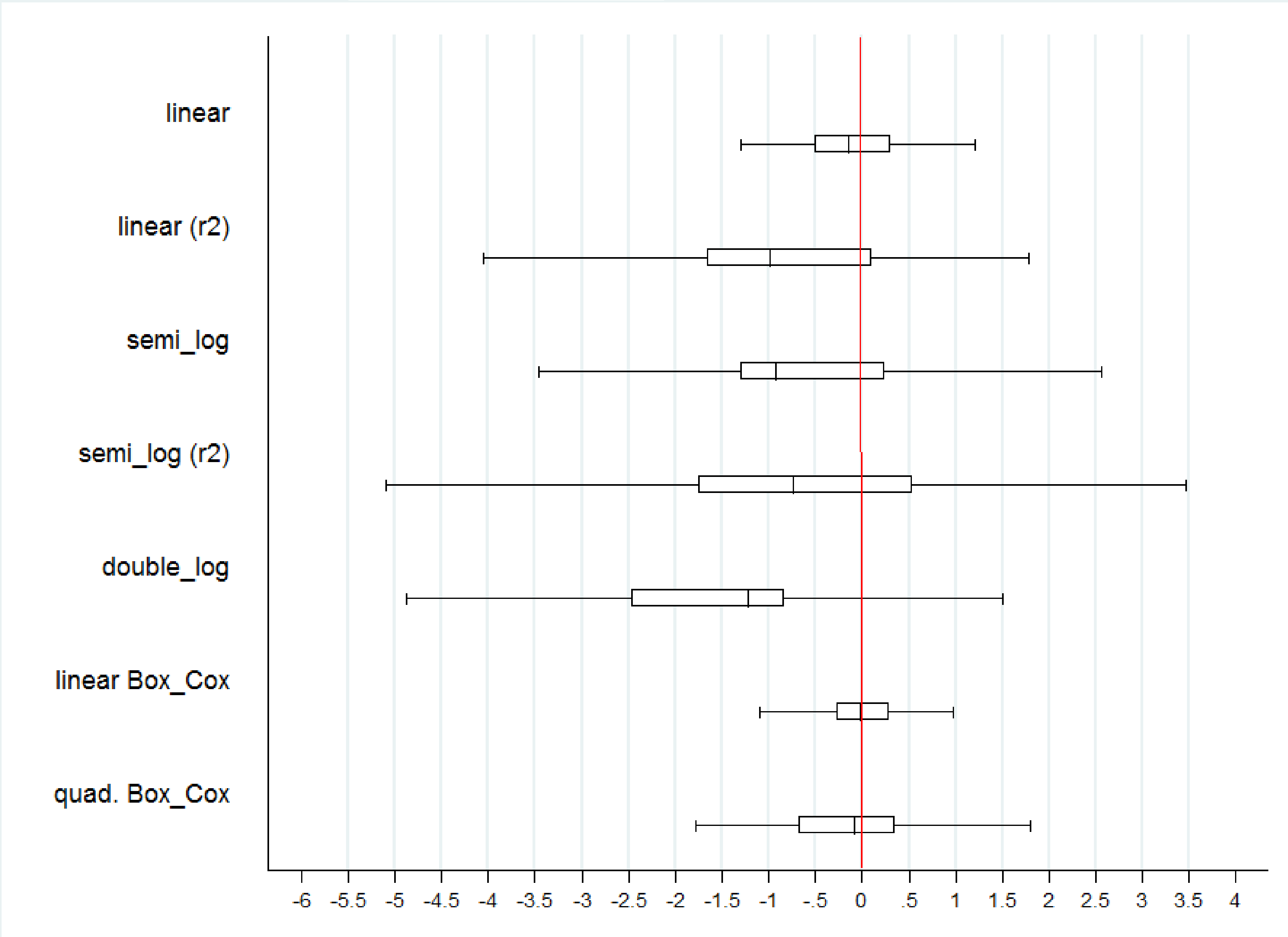
- Rely on meta regression to summarize influence of key model choices:
  - Functional forms
  - Occupation/industry controls
  - β_p estimators
  - Choice of omitted variables

$$|\beta_p| = \alpha + \sum_{i=1}^{6} \beta_i \, functional\_\text{form} + \sum_{j=1}^{3} \varphi_j \, occ./ind\_fixed\_effect + \text{exposure} +$$

$$\text{cognitive\_skill} + \text{motor\_skill} + \text{inter\_people} + \text{physical demand} + \varepsilon$$

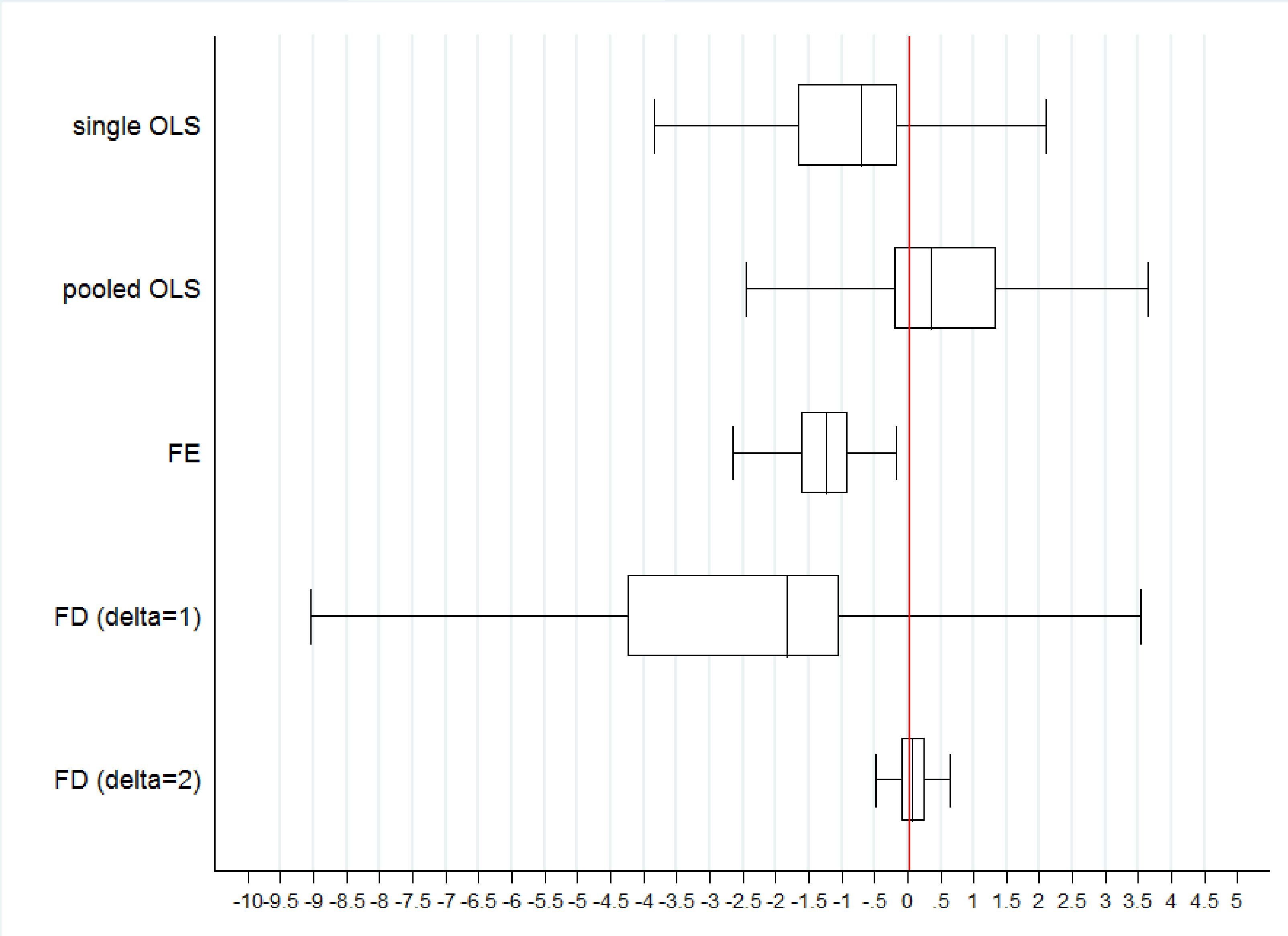
Distribution of biases across functional forms

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \; , \; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



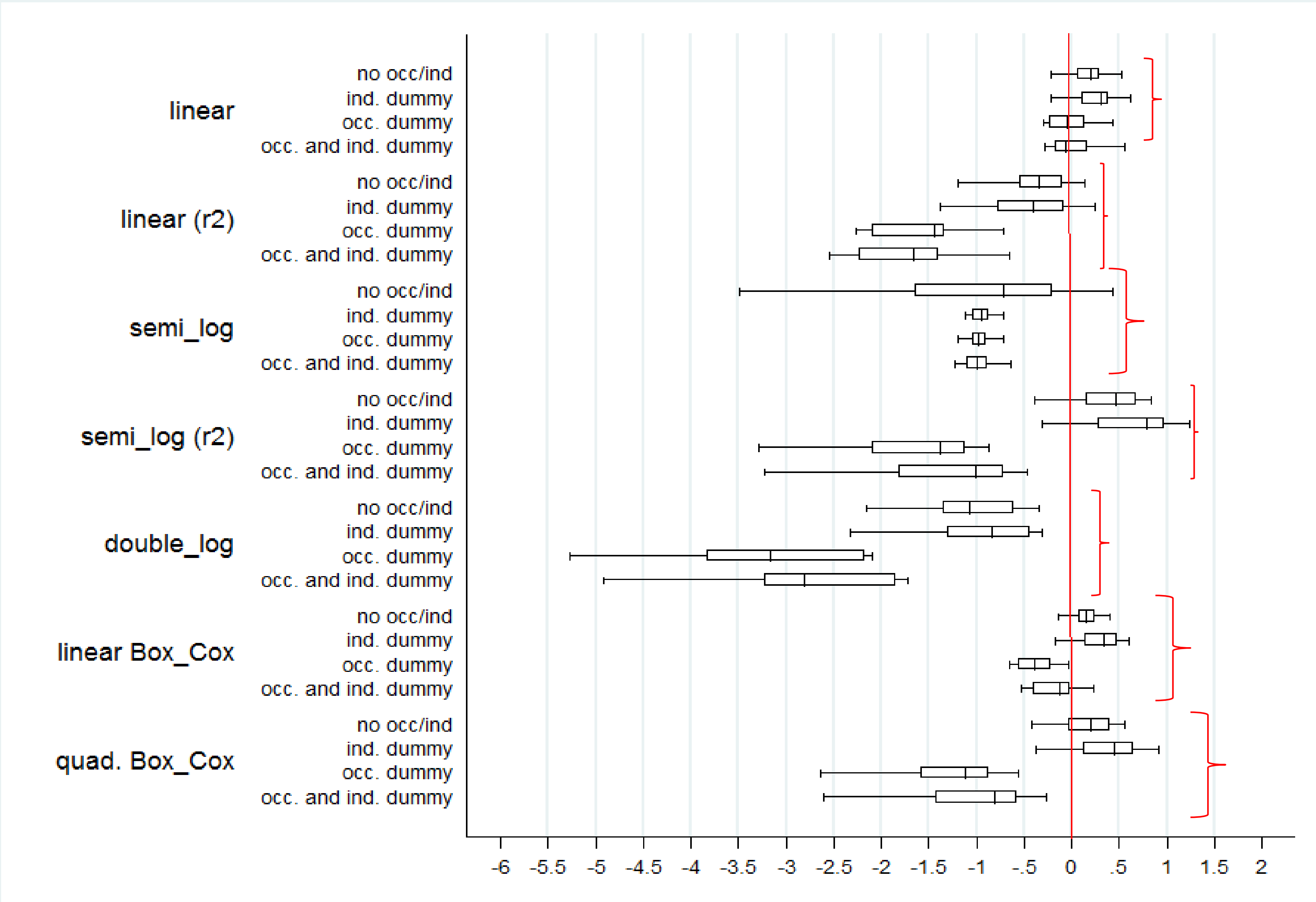
Distribution of biases across estimators

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \; , \; \beta_p = \frac{\bar{e}_p}{N^{-1} \sum_i MWTP_i}, \quad \right)$$



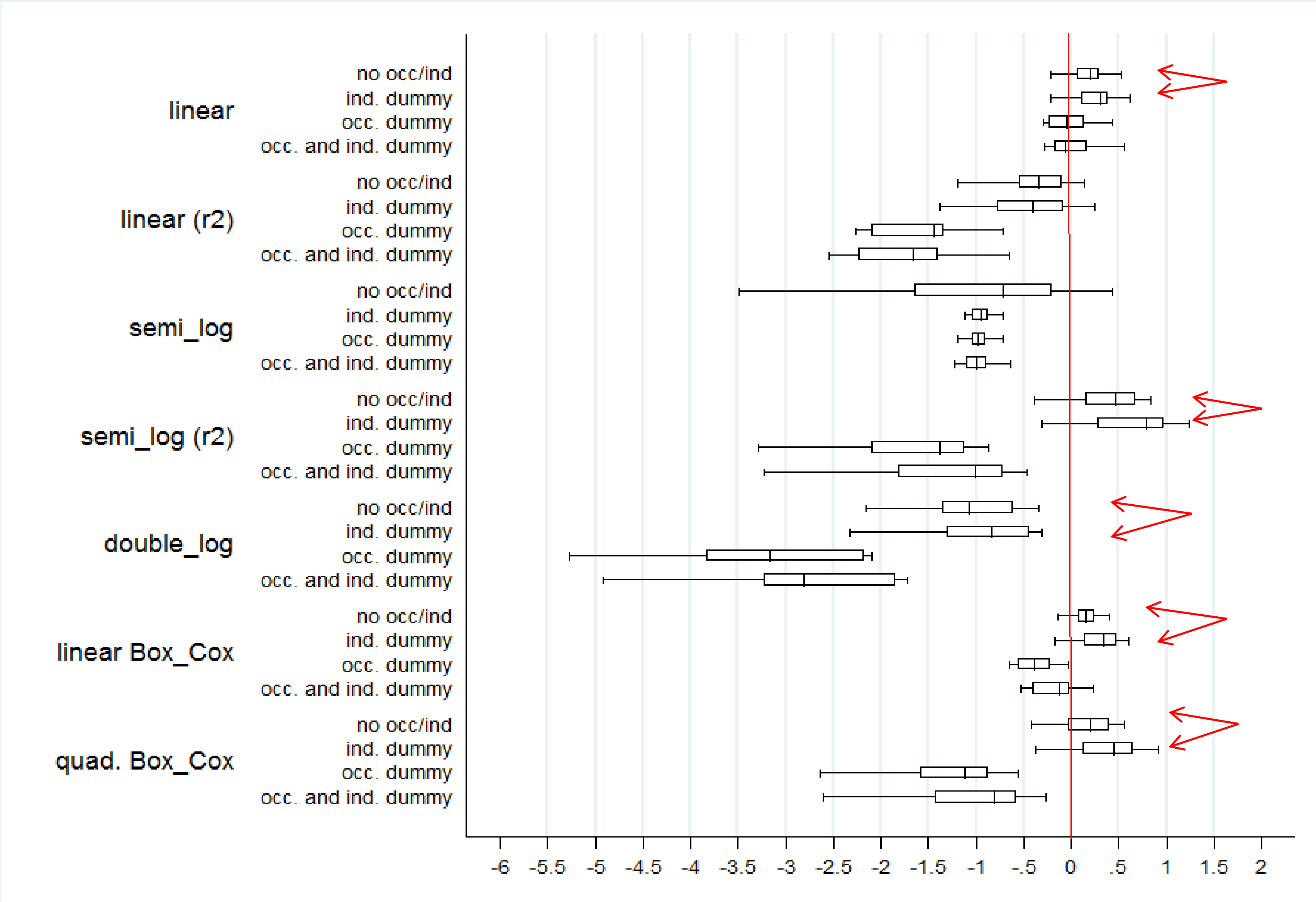
Distribution of biases

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \,,\ \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



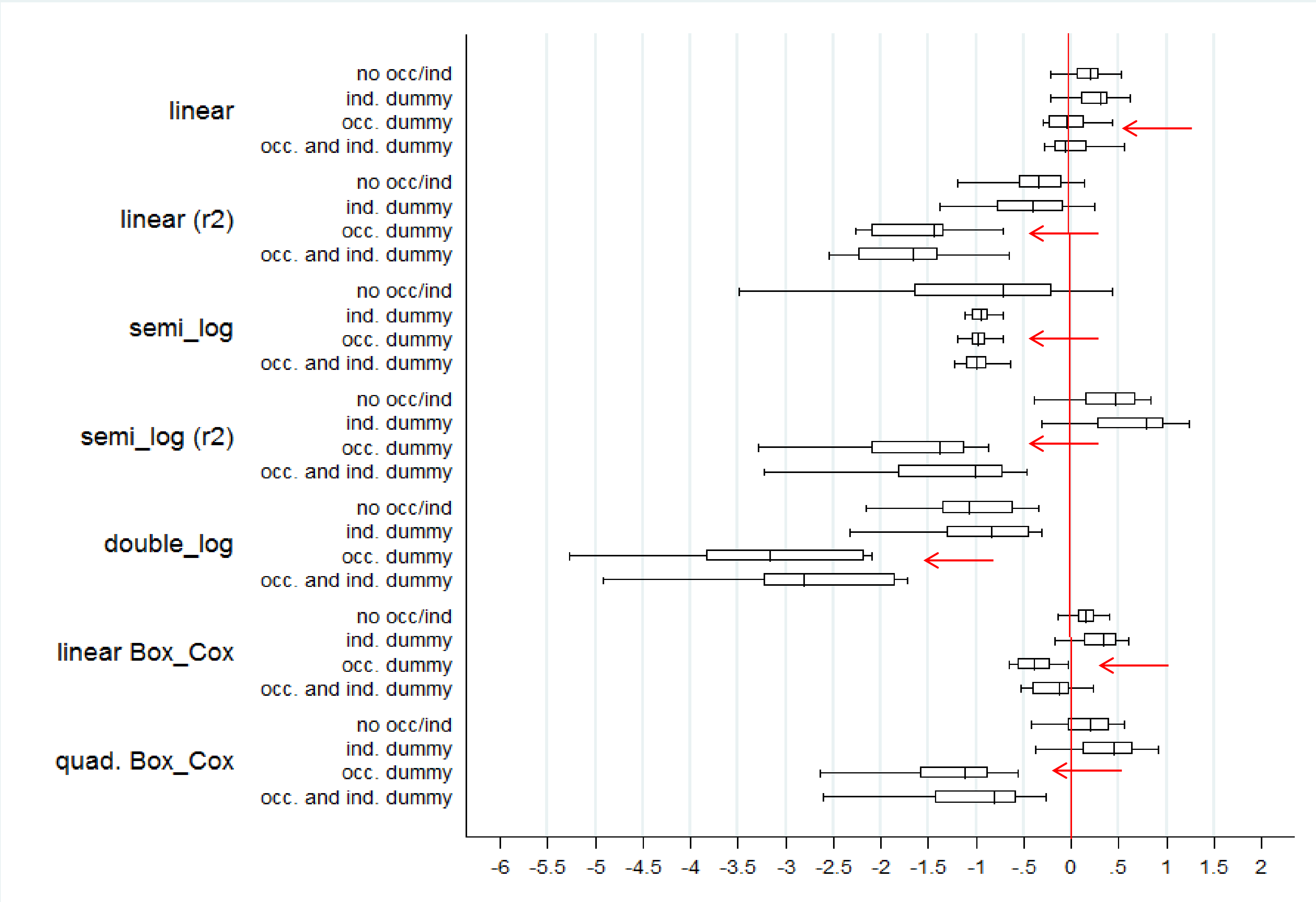
Distribution of biases

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \ , \ \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



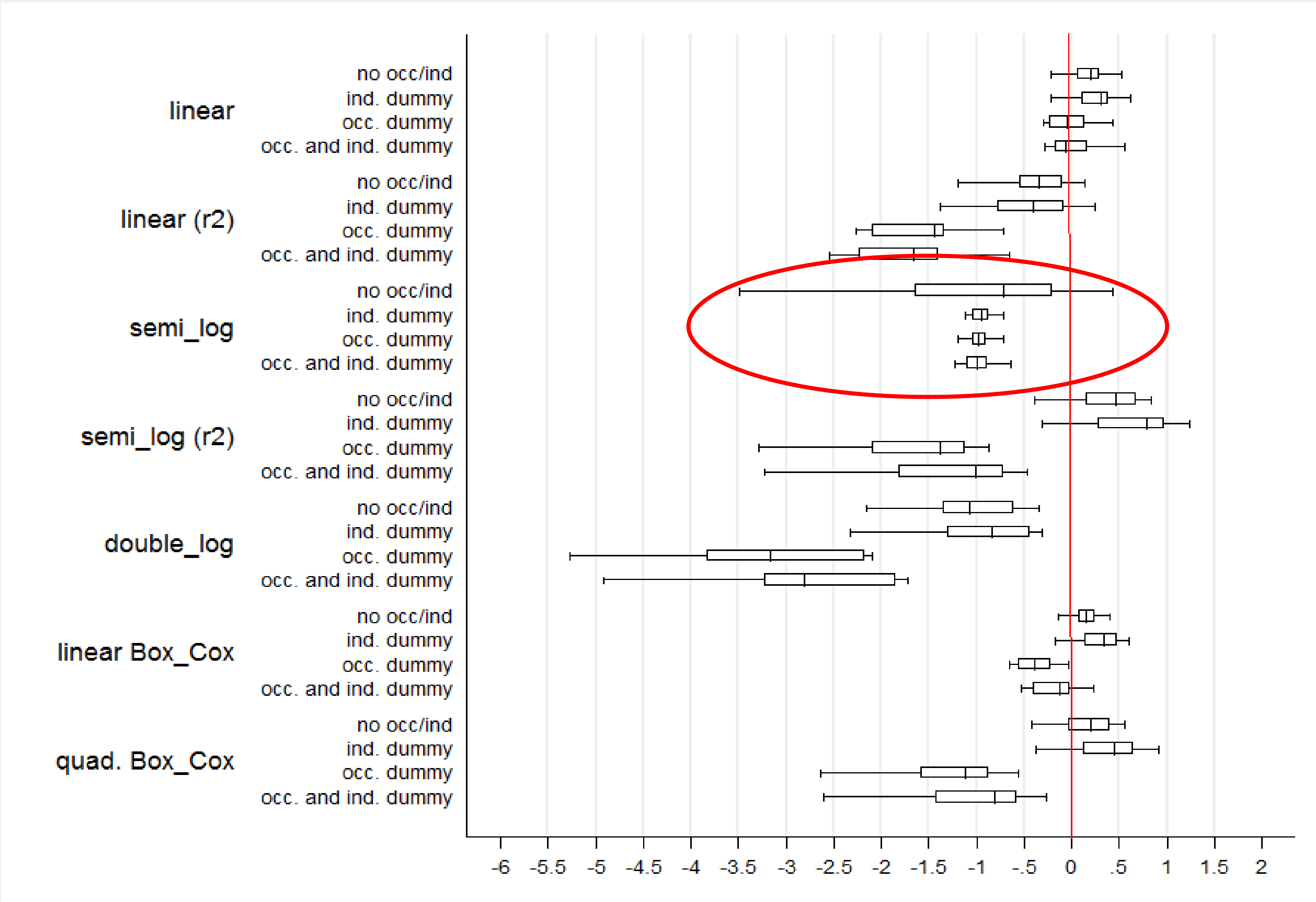
Distribution of biases

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \ , \ \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



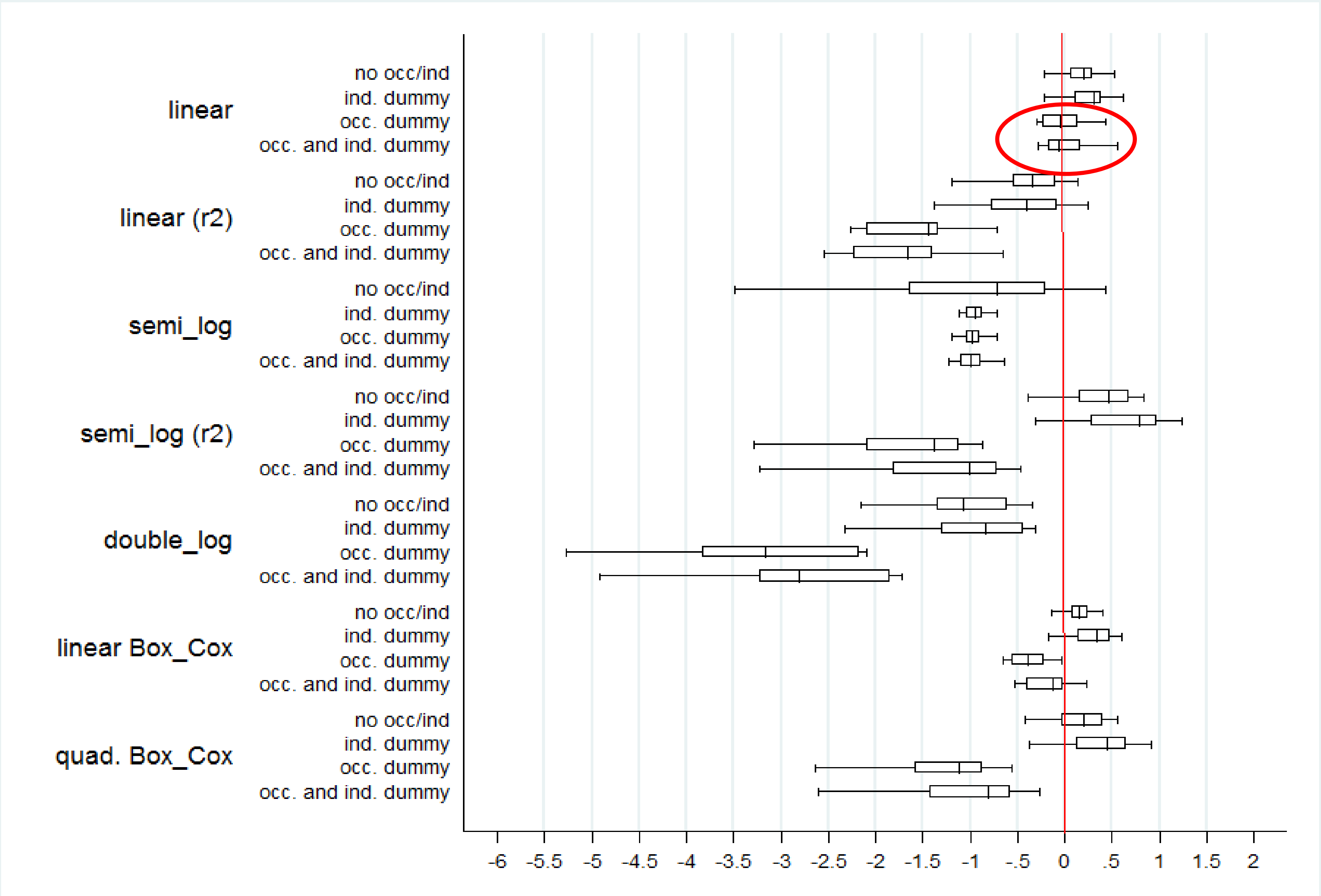
Distribution of biases

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \,,\; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



Distribution of biases

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \;,\; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



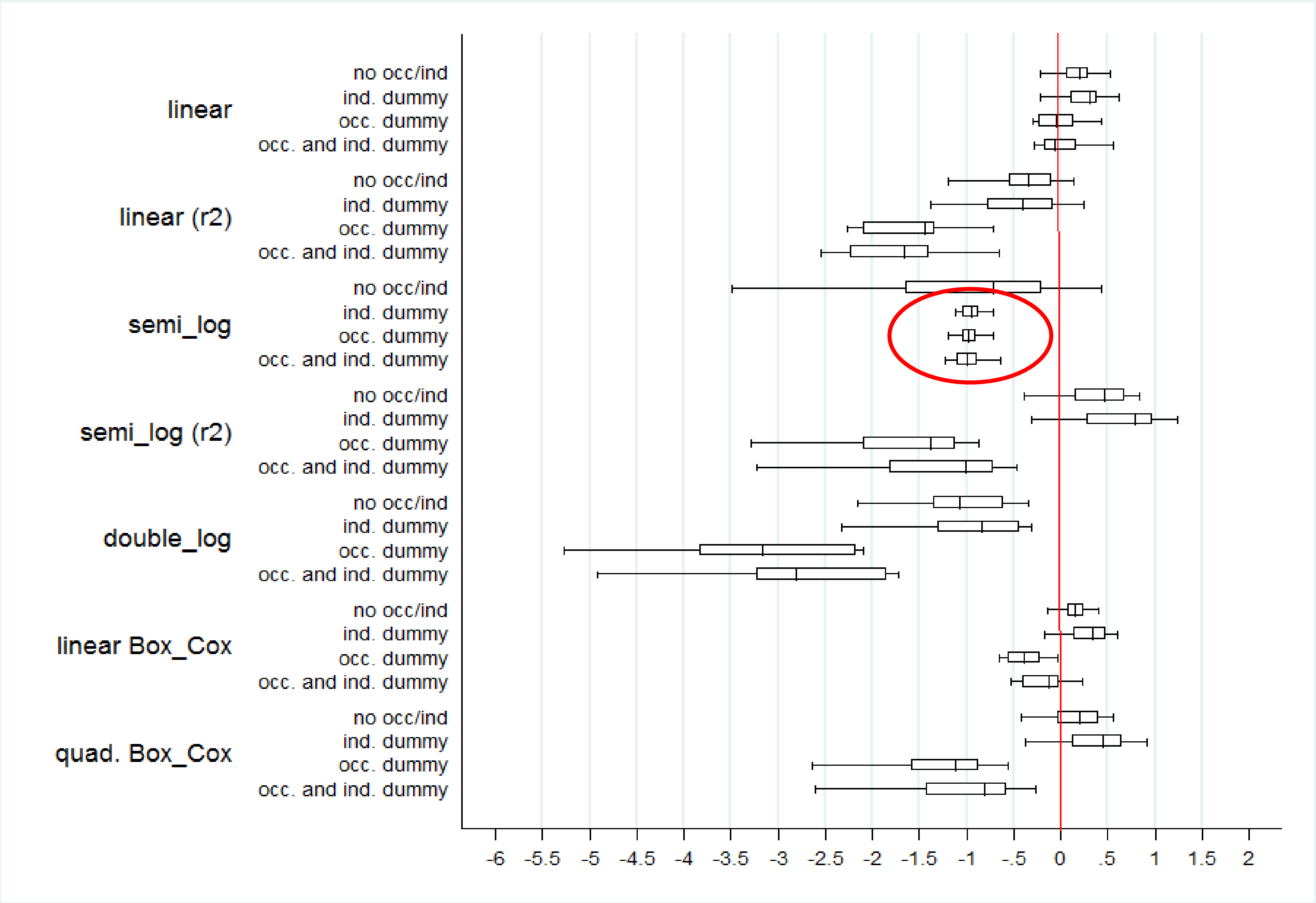
A simulation approach to evaluate the modeling specifications

- Which functional form is appropriate?
  - Linear and linear Box-Cox functional forms are less vulnerable to omitted variables.
- Which estimator is more preferable?
  - FD (Δ=2) tends to produce least bias and is more robust to the model specifications.
- Does occupation and/or industry fixed effects help to increase the accuracy of the estimates?
  - Adding occupation fixed effects improves models' performance when it is applied to linear regression model.

Thank you.

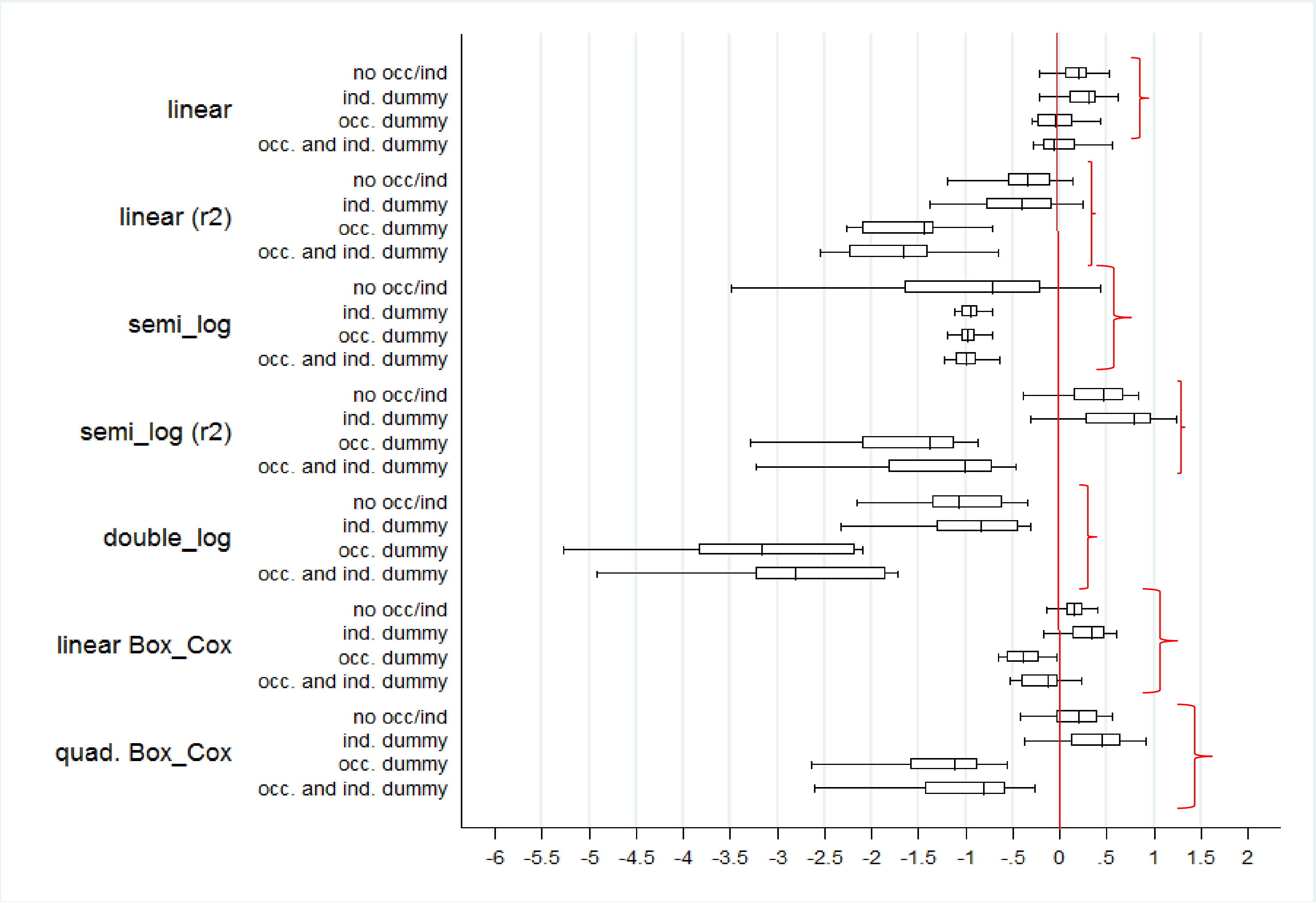
Distribution of biases (*single OLS*)

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \; , \; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



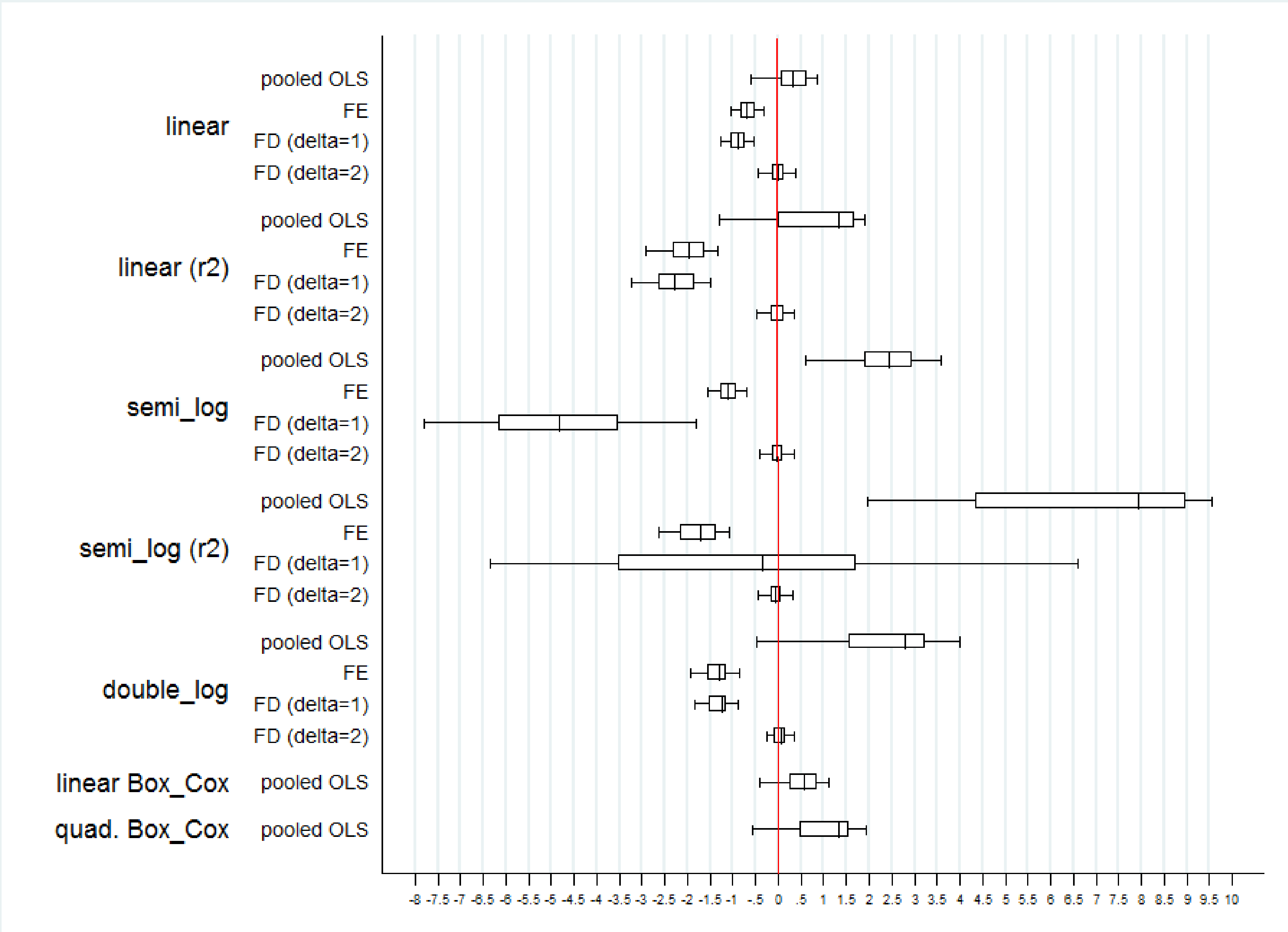
Distribution of biases (*single OLS*)

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \,,\ \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



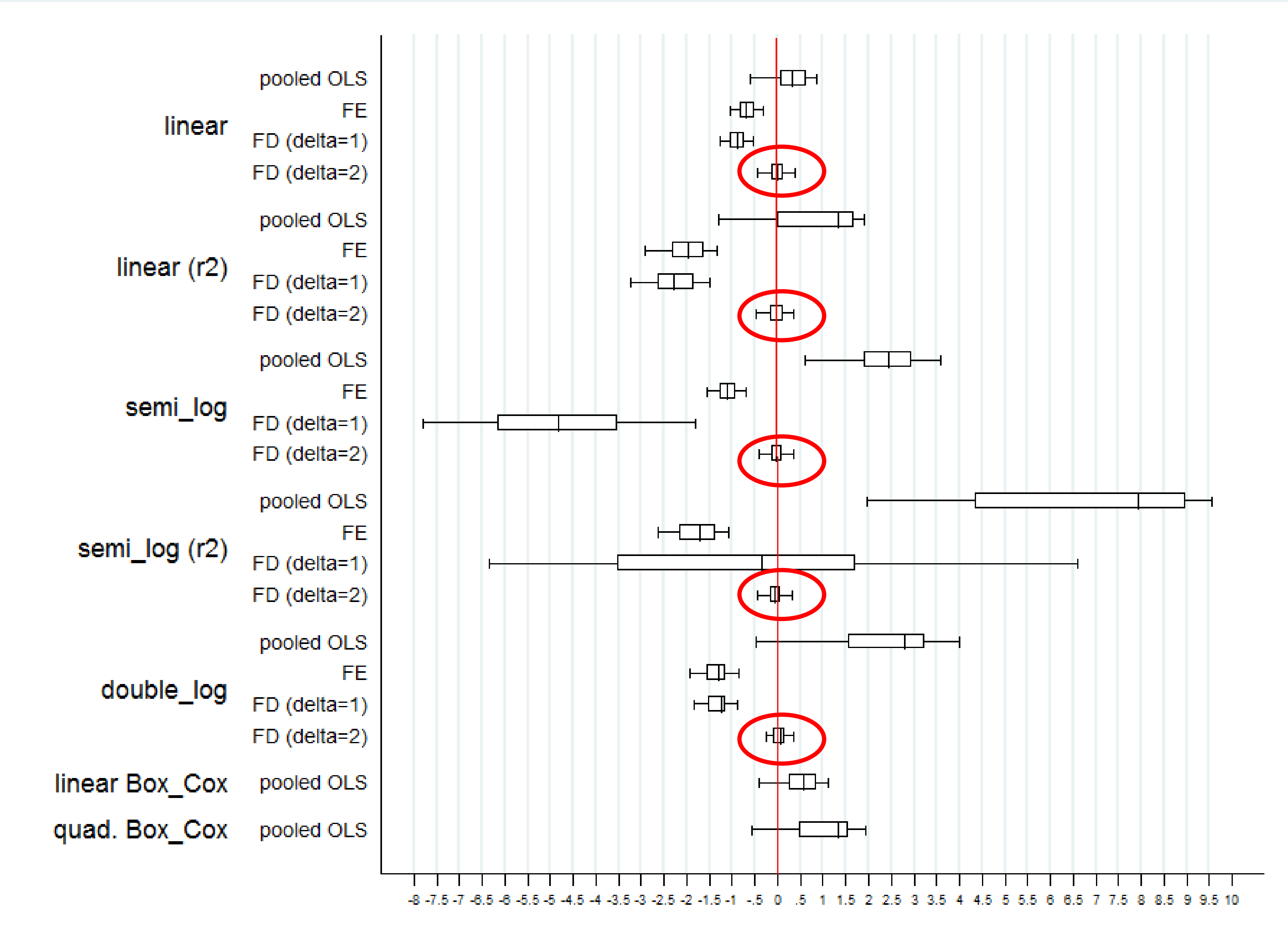
Distribution of biases (no occ./ind.)

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \; , \; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



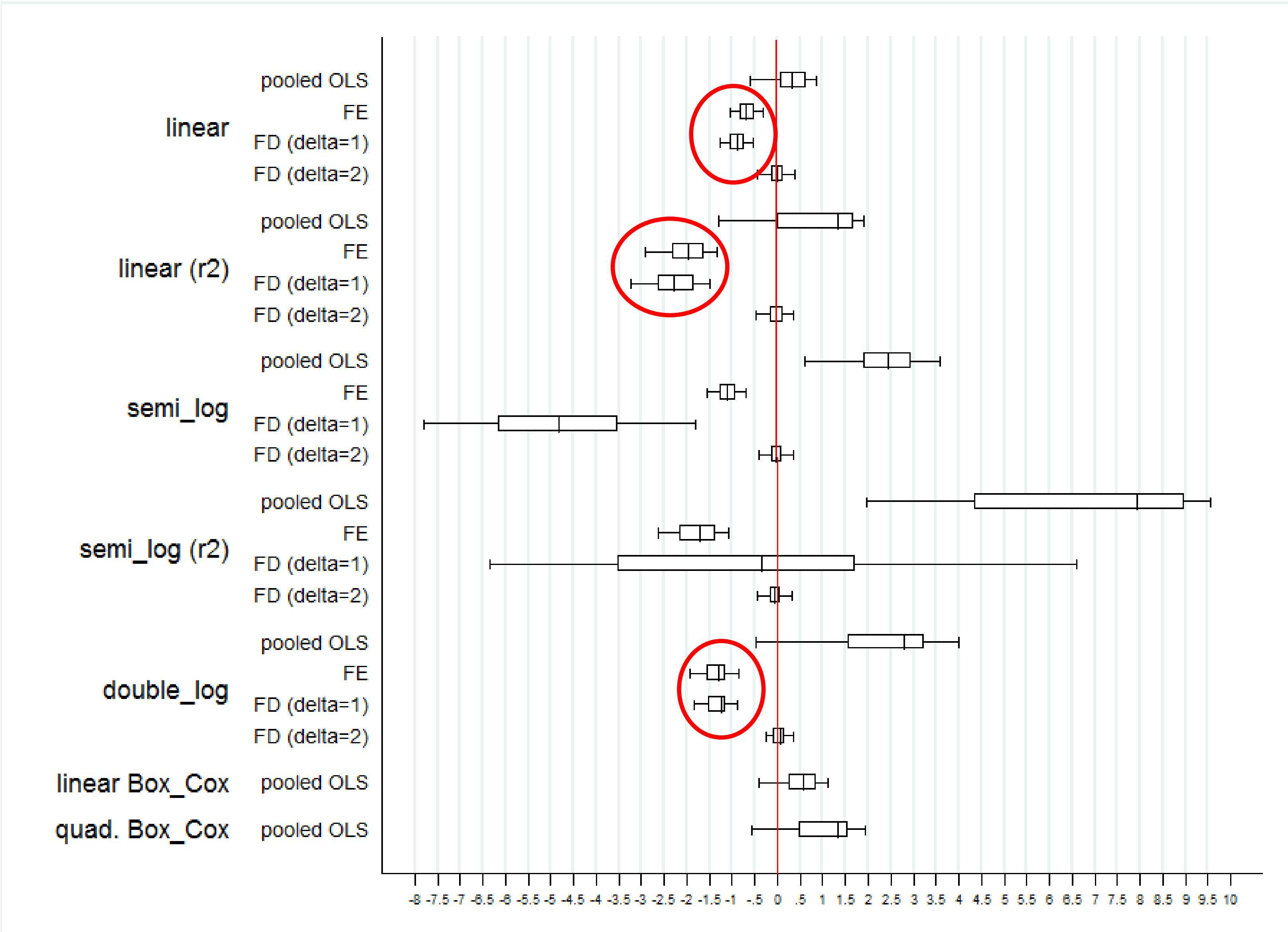
Distribution of biases (no occ./ind.)

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \;,\; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



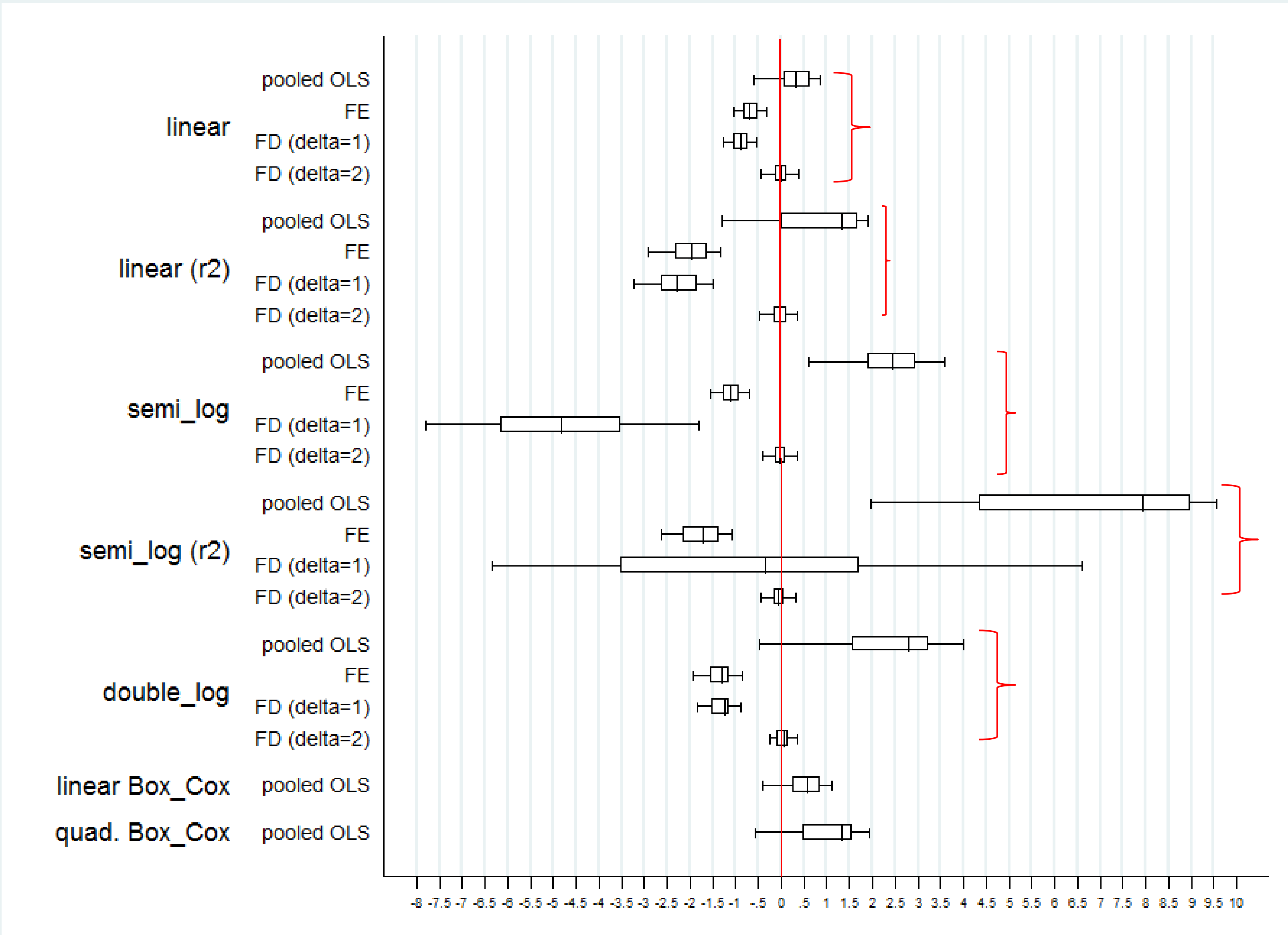
Distribution of biases (no occ./ind.)

$$\left(\quad e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \;,\; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \quad \right)$$



Distribution of biases (no occ./ind.)

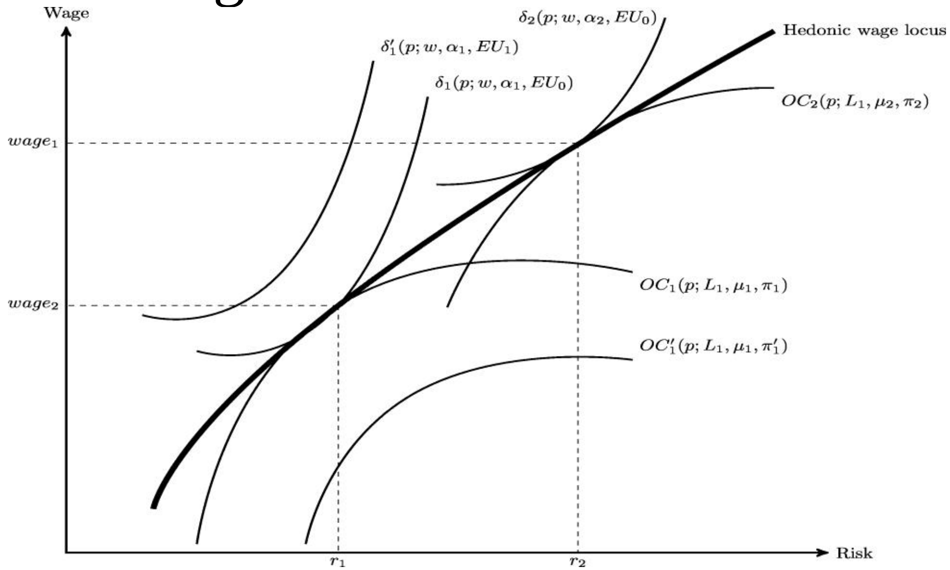
$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \,,\; \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



Thank you.

A simulation approach to evaluate the modeling specifications

- Hedonic wage theory
  - Hedonic wage curve:



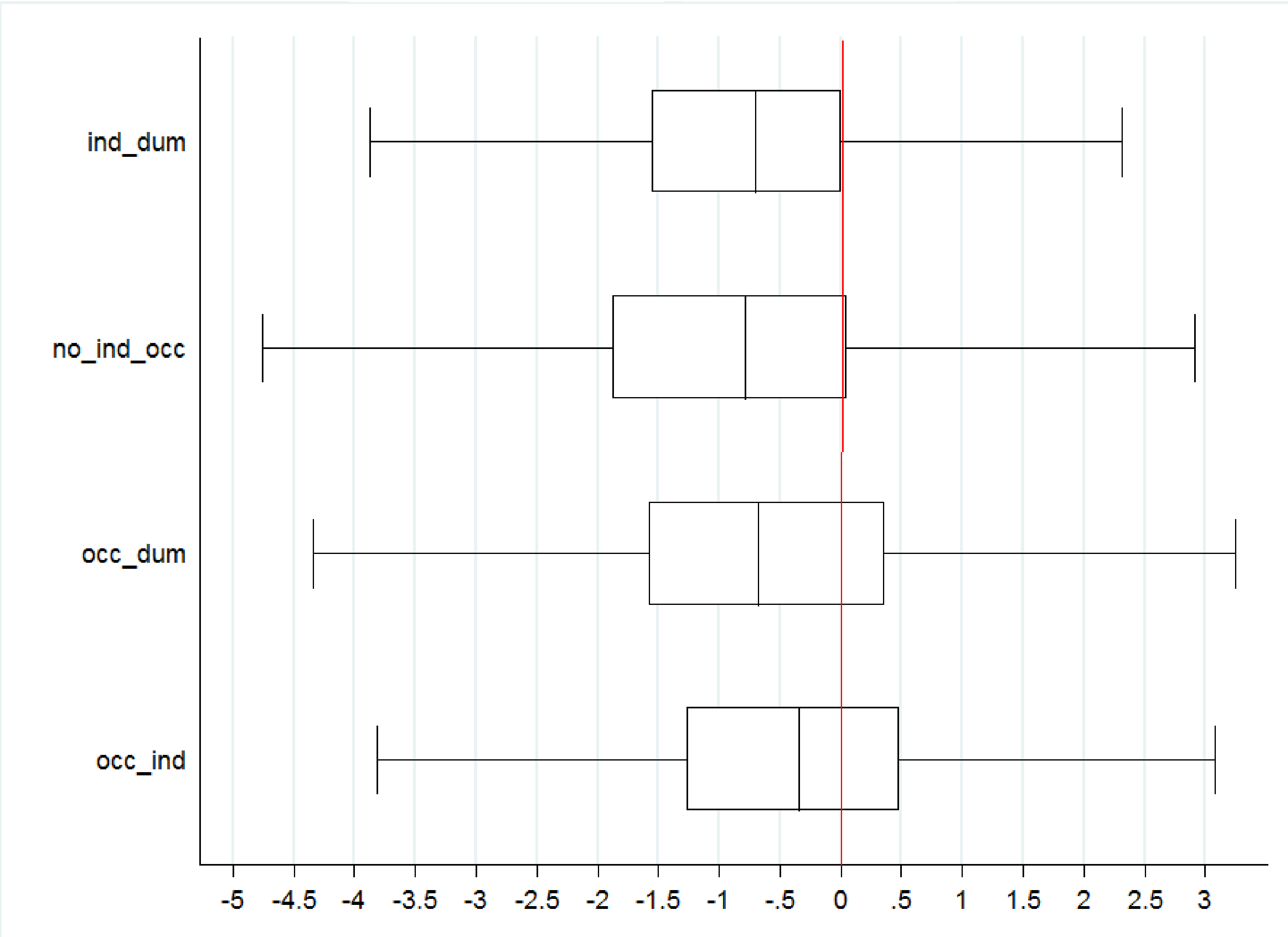
  - "True" marginal-willingness-to pay (MWTP):

$$E(U)=\max\ (1-p)\, U_0\,(\bar{y} + W(p)) + p\, U_1\,(\bar{y} + \widetilde{W}),$$

$$\text{MWTP} \equiv \frac{dw}{dp} = \frac{U_0\big(\bar{y} + W(p)\big) - \ U_1(\bar{y} + \widetilde{W})}{(1-p)U_0'\big(\bar{y} + W(p)\big) + \ pU_1'(\bar{y} + \widetilde{W})},$$

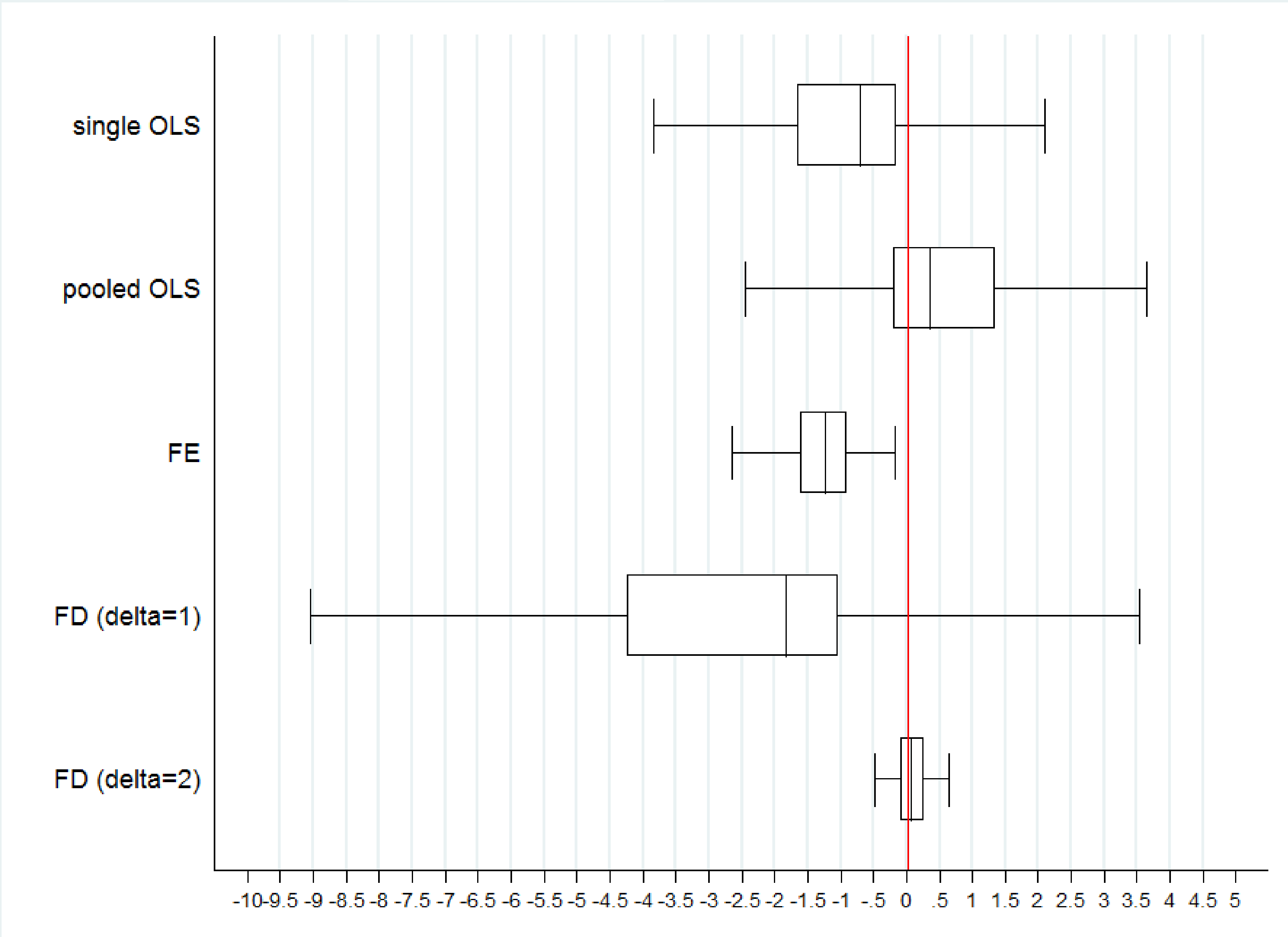
Distribution of biases across occ./ind. fixed effect

($e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i$, $\beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}$,)



Distribution of biases across estimators

$$\left(e_{pi} = \frac{\partial W(p_i)}{\partial p_i} - MWTP_i \,,\ \beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \right)$$



Stage 2: Evaluating hedonic wage model specifications

- Evaluation method
  - Summary statistics of bias of each hedonic wage model:

$$\beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i} \qquad S_p = \frac{\bar{s}_p}{N^{-1}\sum_i MWTP_i}$$

β_p < -1: $N^{-1}\sum_i \frac{\partial W(p_i)}{\partial p_i} < 0$

$\beta_p \cong$ -1: $N^{-1}\sum_i \frac{\partial W(p_i)}{\partial p_i} \cong 0$

$\beta_p \cong$ 1: $N^{-1}\sum_i \frac{\partial W(p_i)}{\partial p_i} \cong$ **2 x** $N^{-1}\sum_i MWTP_i$

$\Rightarrow$ # of β_p and S_p : ***3,480***.