

Misspecification of preference heterogeneity structure in hybrid choice models

Wiktor Budziński



NATIONAL SCIENCE CENTRE
POLAND

Motivation

- Hybrid choice models are becoming a popular method for analyzing data from discrete choice experiments
 - They allow for inclusion of '*soft*' variables such as perceptions and attitudes into the choice model using latent variables framework
 - Considered to be more '*behavioral*' approach for explaining preference heterogeneity
- A lot of effort put into understanding preference heterogeneity in standard mixed logit model
- In contrast, limited effort put to understand analogous issues in hybrid choice models
 - We investigate these issues using Monte Carlo simulation

Hybrid choice models

- Based on random utility model

$$U_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}_i + \varepsilon_{ij}$$

- Individual-specific parameters can be decomposed:

$$\boldsymbol{\beta}_i = \alpha\mathbf{SD}_i + \lambda\mathbf{LV}_i + \boldsymbol{\beta}_i^*$$

- Furthermore, latent variables can be decomposed as

$$\mathbf{LV}_i = \gamma\mathbf{SD}_i + \xi_i$$

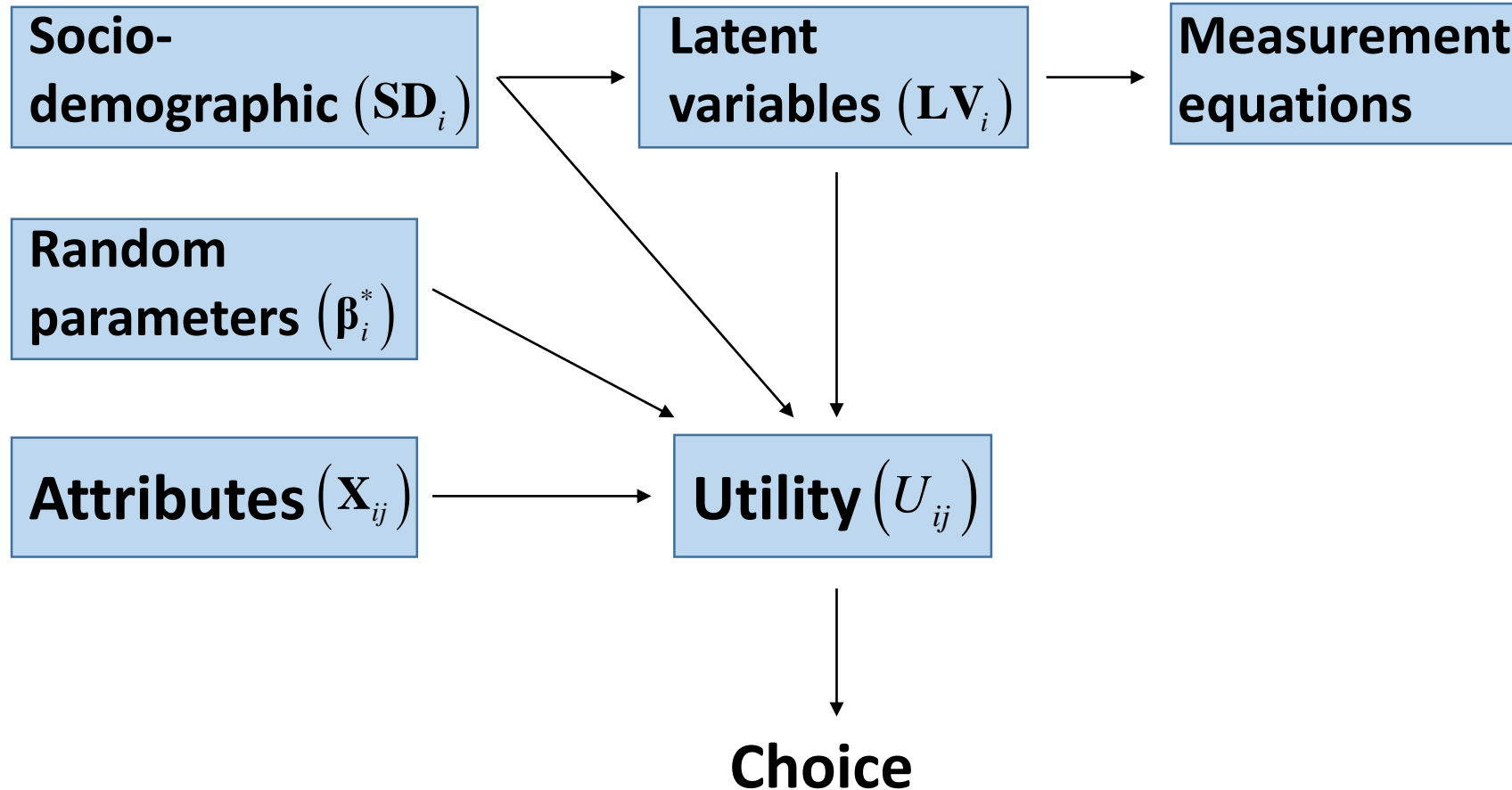
Hybrid choice models

- In contrast, in standard mixed logit model we have

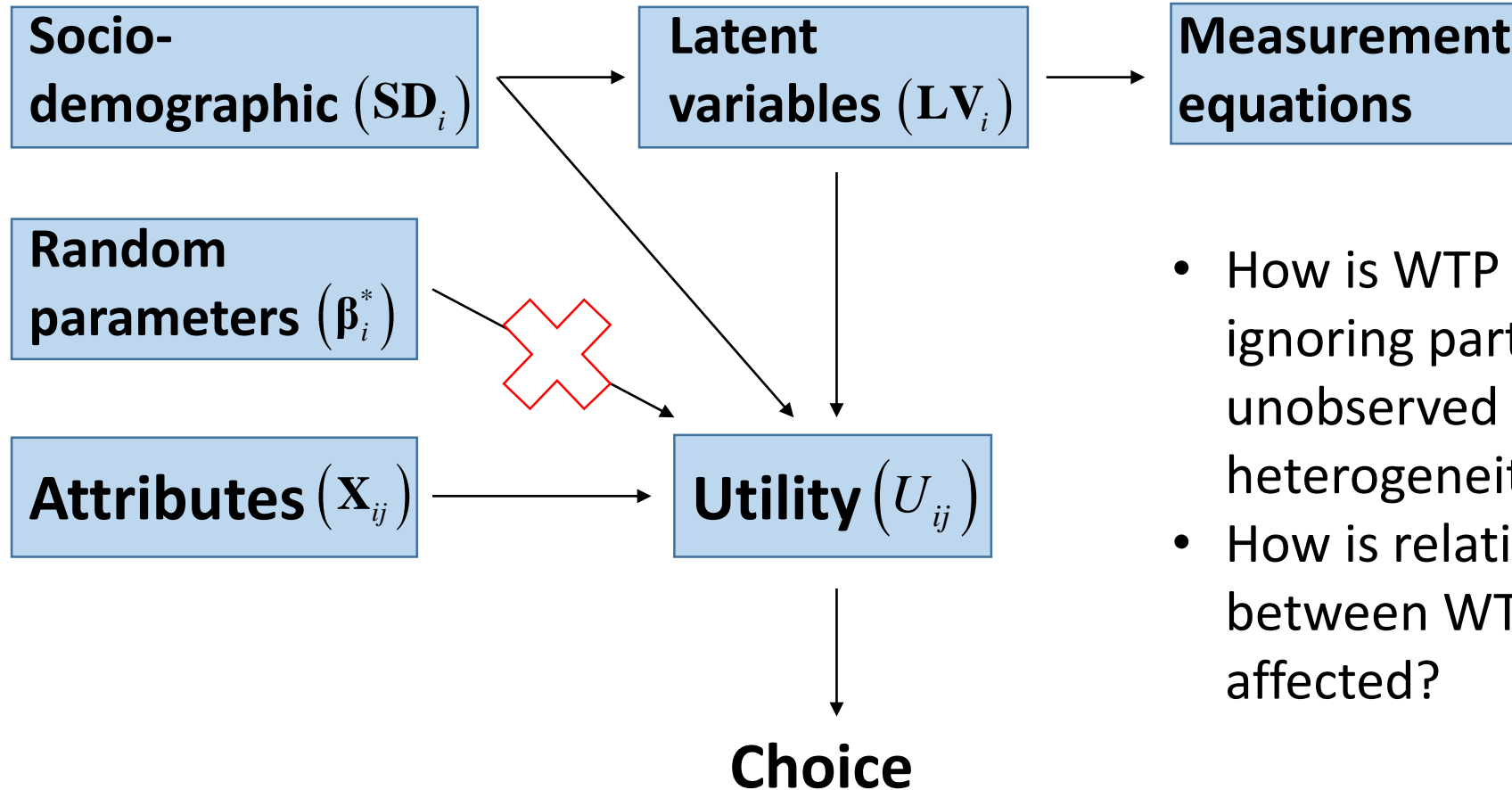
$$\beta_i = \alpha^* \mathbf{SD}_i + \eta_i^*$$

- Mixed logit can be therefore seen as a *reduced form model* (Vij and Walker, 2016)
 - Latent variables get incorporated into other sources of preference heterogeneity
 - Can be used to guide specification of hybrid choice models
- We should not expect to obtain much “*better*” results from hybrid models in general, but:
 - We gain better understanding of human behavior
 - Estimates can be more precise, as hybrid models use additional data

Hybrid choice models – general framework

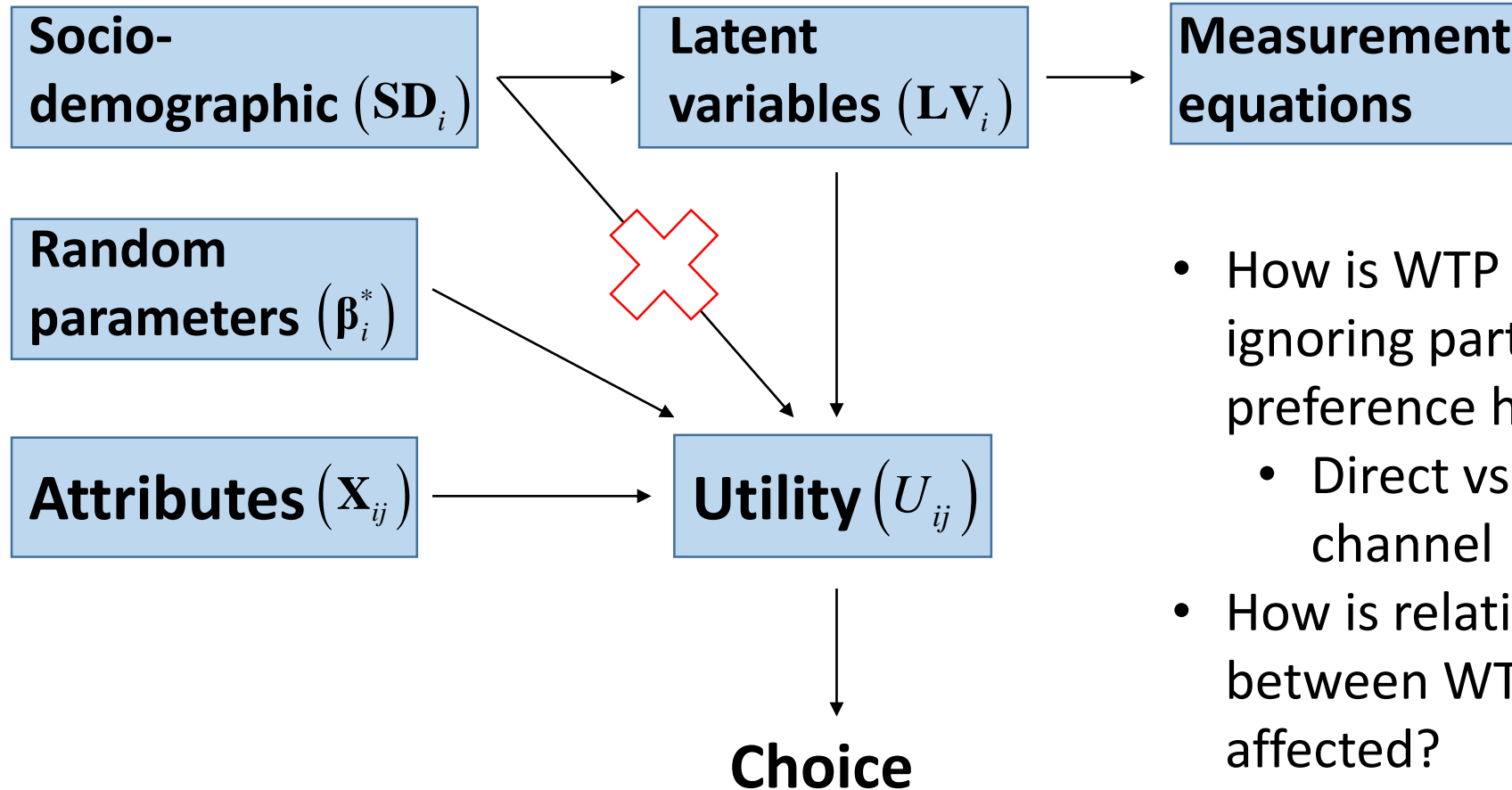


Hybrid choice models – simplification of unobserved heterogeneity



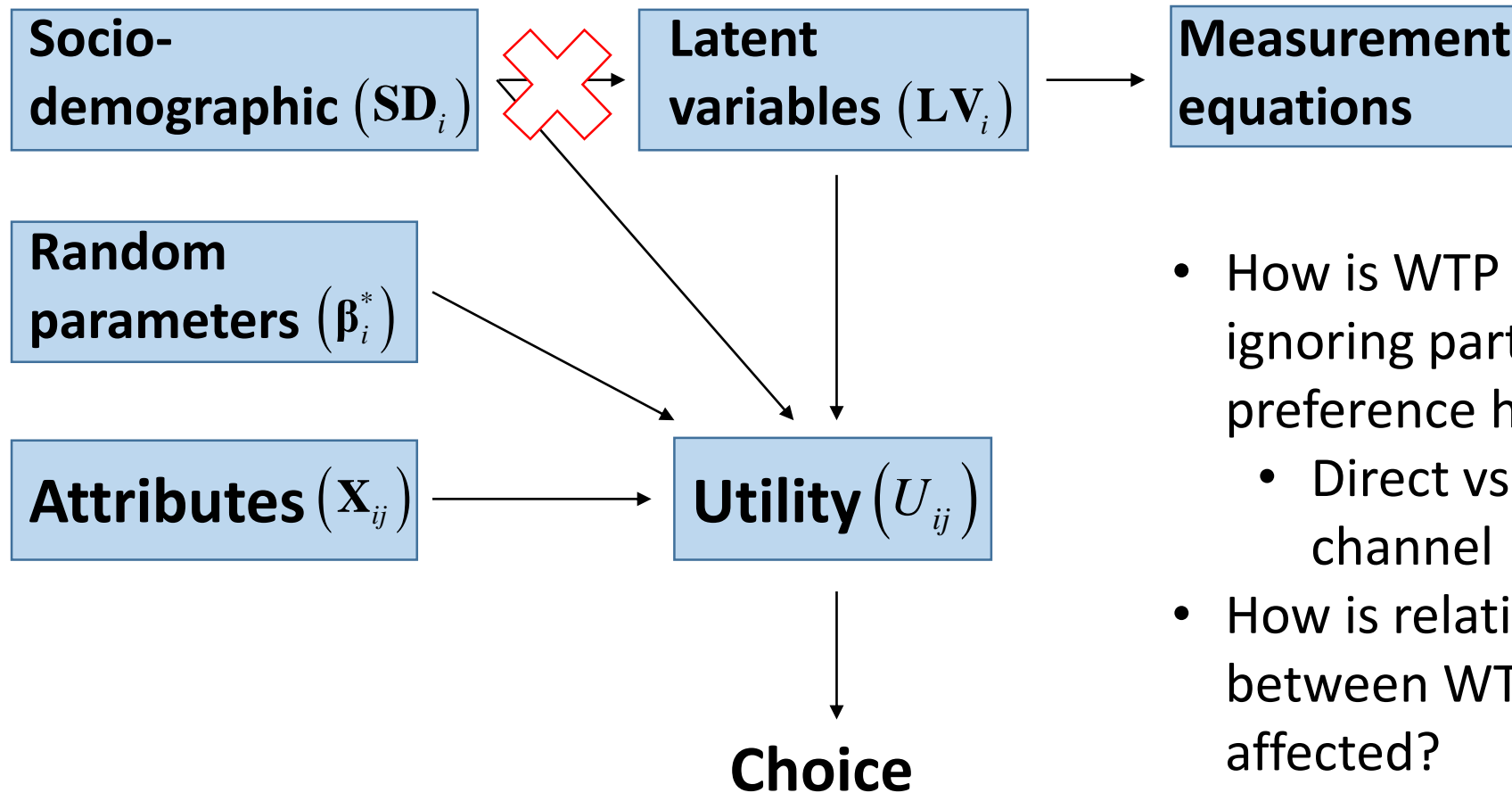
- How is WTP affected by ignoring part of unobserved preference heterogeneity?
- How is relationship between WTP and LV affected?

Hybrid choice models – simplification of observed heterogeneity



- How is WTP affected by ignoring part of observed preference heterogeneity?
 - Direct vs. indirect channel
- How is relationship between WTP and LV affected?

Hybrid choice models – simplification of observed heterogeneity



- How is WTP affected by ignoring part of observed preference heterogeneity?
 - Direct vs. indirect channel
- How is relationship between WTP and LV affected?

Simulation setting

- Simulation mimics a simple stated preference study
 - 3 alternatives (one *status quo*)
 - 3 attributes ($Quality_1$, $Quality_2$, $Cost$) + status quo ASC
 - 10 choice tasks per individual
 - 2000 individuals
- Separate data generating processes to investigate
 - Misspecification of unobserved heterogeneity
 - Misspecification of observed heterogeneity
- We generate 1000 artificial datasets for each DGP
 - We estimate correct model, reduced form model, and misspecified model(s)

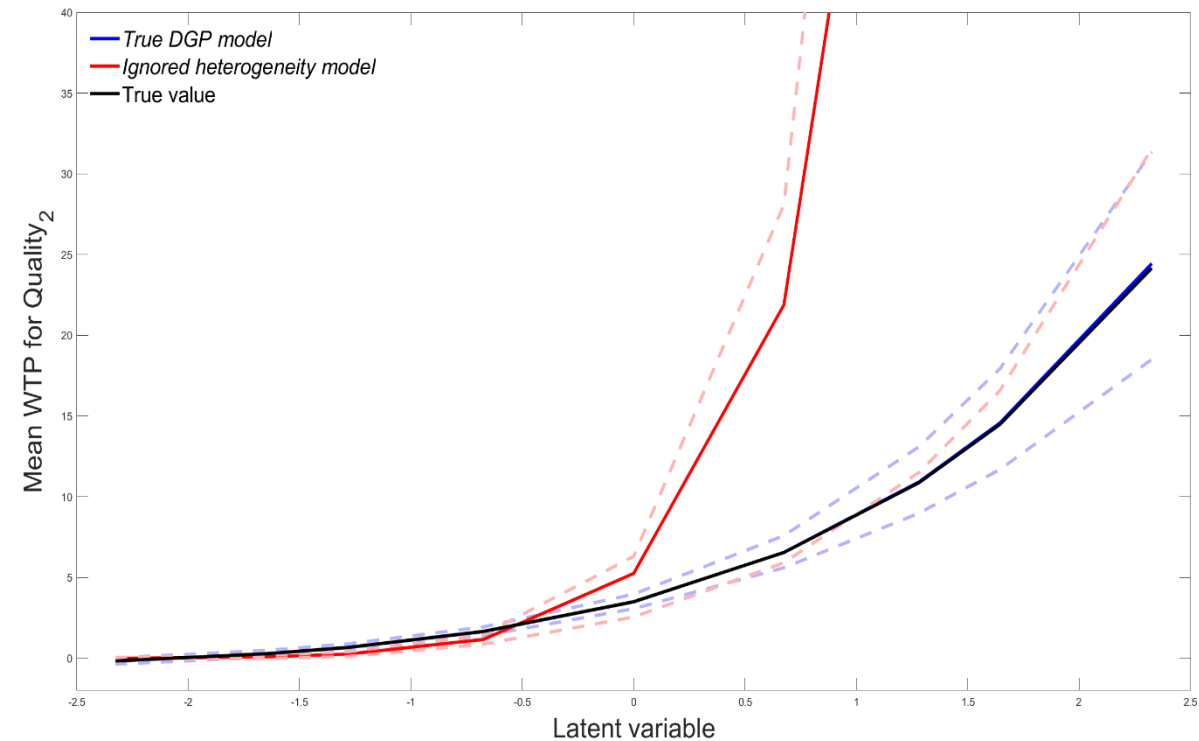
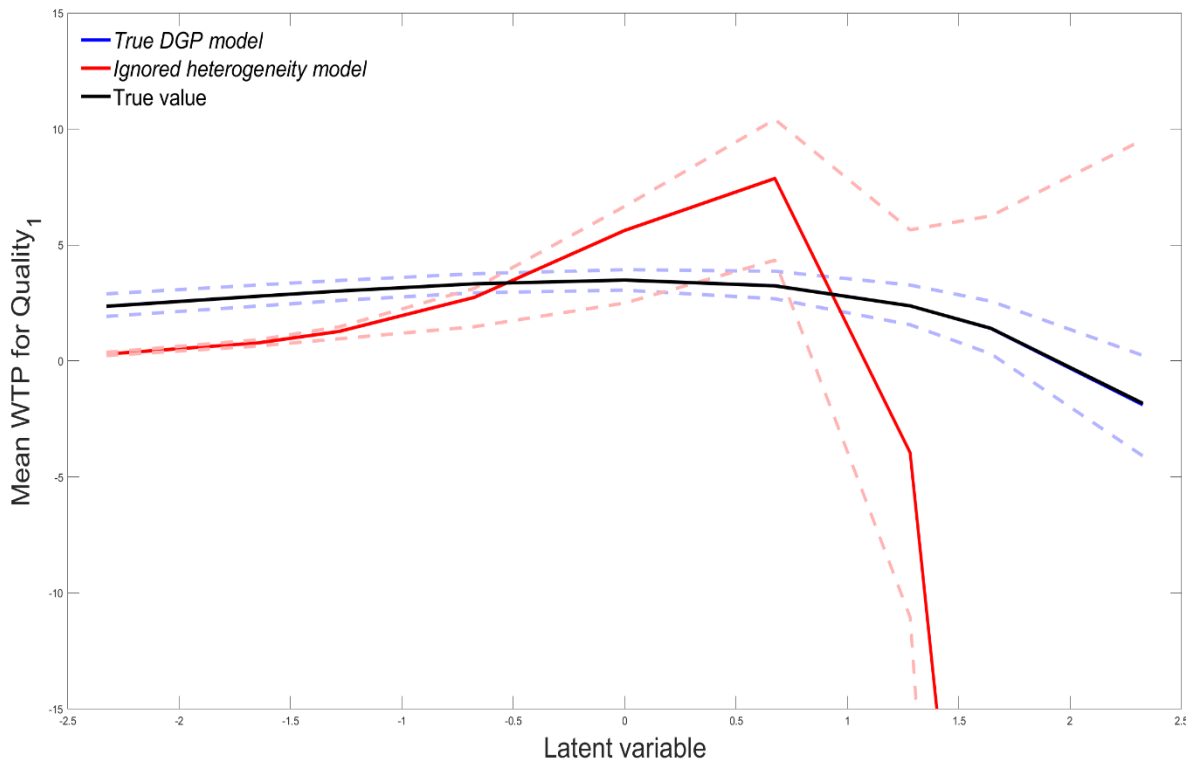
Results – unobserved heterogeneity

- Correct model and reduced form model (MXL) recover WTP well
- Ignoring heterogeneity leads to large bias, especially for mean WTP

	Mean WTP			<i>True value</i>
	<i>True DGP</i>	<i>Ignored heterogeneity</i>	<i>Reduced form model</i>	
<i>SQ</i>	-2.9544*** [-3.4621, -2.4955]	30.8126* [-3.5063, 53.8059]	-2.9458** [-3.5976, -2.2415]	-2.9668
<i>Quality₁</i>	2.9652*** [2.4935, 3.4982]	-11.9039 [-23.8744, 2.9582]	2.9805** [2.2761, 3.7823]	2.9678
<i>Quality₂</i>	4.9566*** [4.1006, 6.0484]	42.3574* [3.2933, 68.5832]	4.9445*** [3.8997, 6.2625]	4.9415
	Median WTP			
	<i>True DGP</i>	<i>Ignored heterogeneity</i>	<i>Reduced form model</i>	<i>True value</i>
<i>SQ</i>	-2.0692*** [-2.2836, -1.8824]	-2.7286 [-3.1132, -2.3792]	-2.0743*** [-2.2906, -1.8713]	-2.0738
<i>Quality₁</i>	2.0749*** [1.8624, 2.3019]	3.9127 [2.2644, 4.7639]	2.0779*** [1.8608, 2.3204]	2.0724
<i>Quality₂</i>	2.2861*** [2.0321, 2.5635]	5.2423 [2.2713, 6.4127]	2.2873*** [2.0255, 2.5892]	2.2886

Results – unobserved heterogeneity

- Relation between WTP and latent variable is also biased, when unobserved heterogeneity is ignored



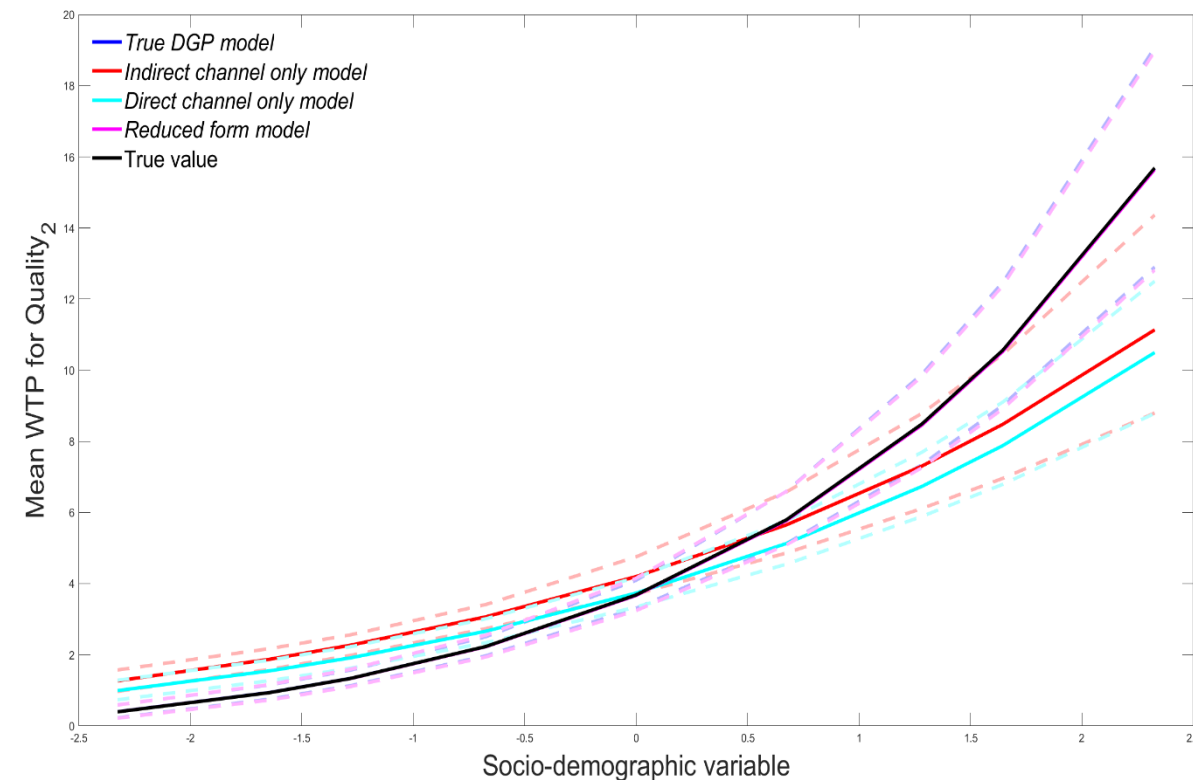
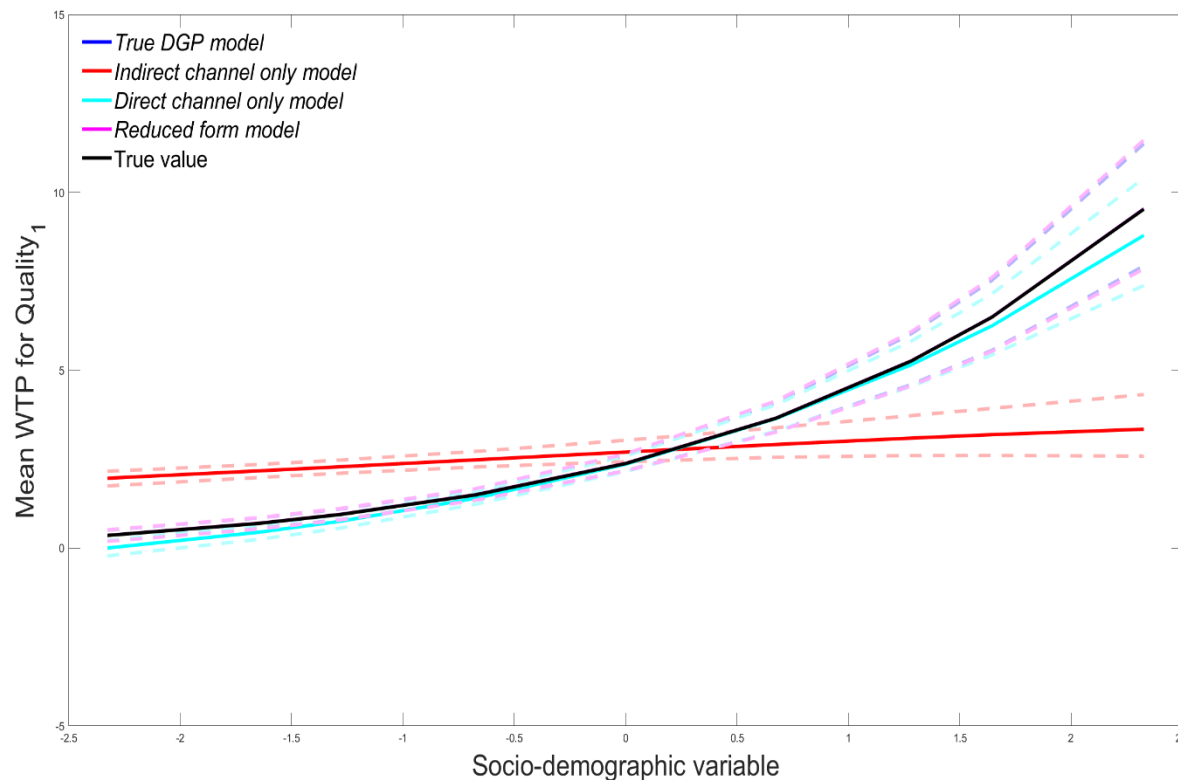
Results – observed heterogeneity

- Much lower bias when compared with misspecification of unobserved heterogeneity

	Mean WTP				<i>True value</i>
	<i>True DGP</i>	<i>Indirect channel only</i>	<i>Direct channel only</i>	<i>Reduced form model</i>	
<i>SQ</i>	-2.3797*** [-2.5779, -2.1883]	-1.9610 [-2.1977, -1.7225]	-2.3564** [-2.5542, -2.1558]	-2.3795*** [-2.5922, -2.1643]	-2.375
<i>Quality₁</i>	2.3788*** [2.1745, 2.5987]	2.6977 [2.4288, 3.0290]	2.3563** [2.1504, 2.5830]	2.3795*** [2.1692, 2.6447]	2.375
<i>Quality₂</i>	3.6758*** [3.3065, 4.0921]	4.1973 [3.7059, 4.7516]	3.7329** [3.3405, 4.1668]	3.6676*** [3.2390, 4.1513]	3.6801
	Median WTP				
	<i>True DGP</i>	<i>Indirect channel only</i>	<i>Direct channel only</i>	<i>Reduced form model</i>	<i>True value</i>
<i>SQ</i>	-2.5946*** [-2.7662, -2.4391]	-2.6040*** [-2.7588, -2.4556]	-2.5859*** [-2.7547, -2.4321]	-2.5968*** [-2.7725, -2.4369]	-2.5883
<i>Quality₁</i>	2.5971*** [2.4038, 2.8048]	2.6903** [2.4780, 2.9400]	2.5890*** [2.3969, 2.7966]	2.6004*** [2.4106, 2.8215]	2.5883
<i>Quality₂</i>	2.7159*** [2.4879, 2.9598]	2.7037*** [2.4710, 2.9540]	2.7156*** [2.4925, 2.9604]	2.7150*** [2.4839, 2.9806]	2.7183

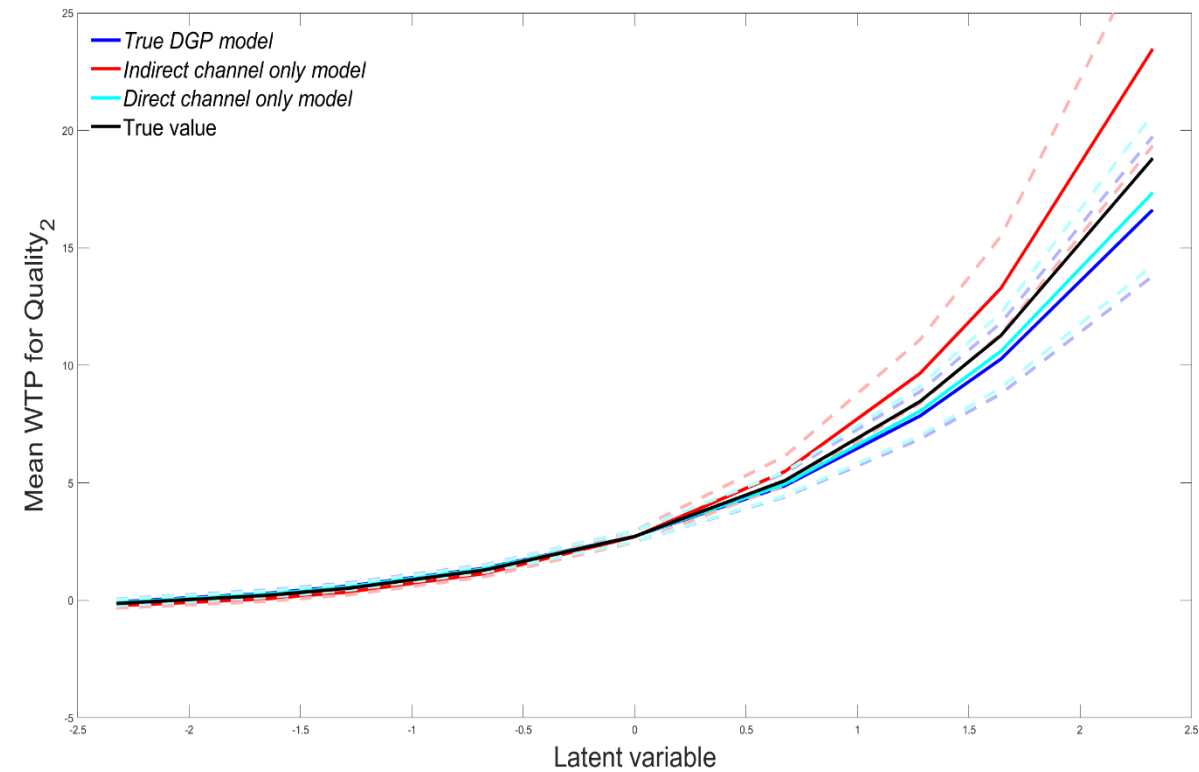
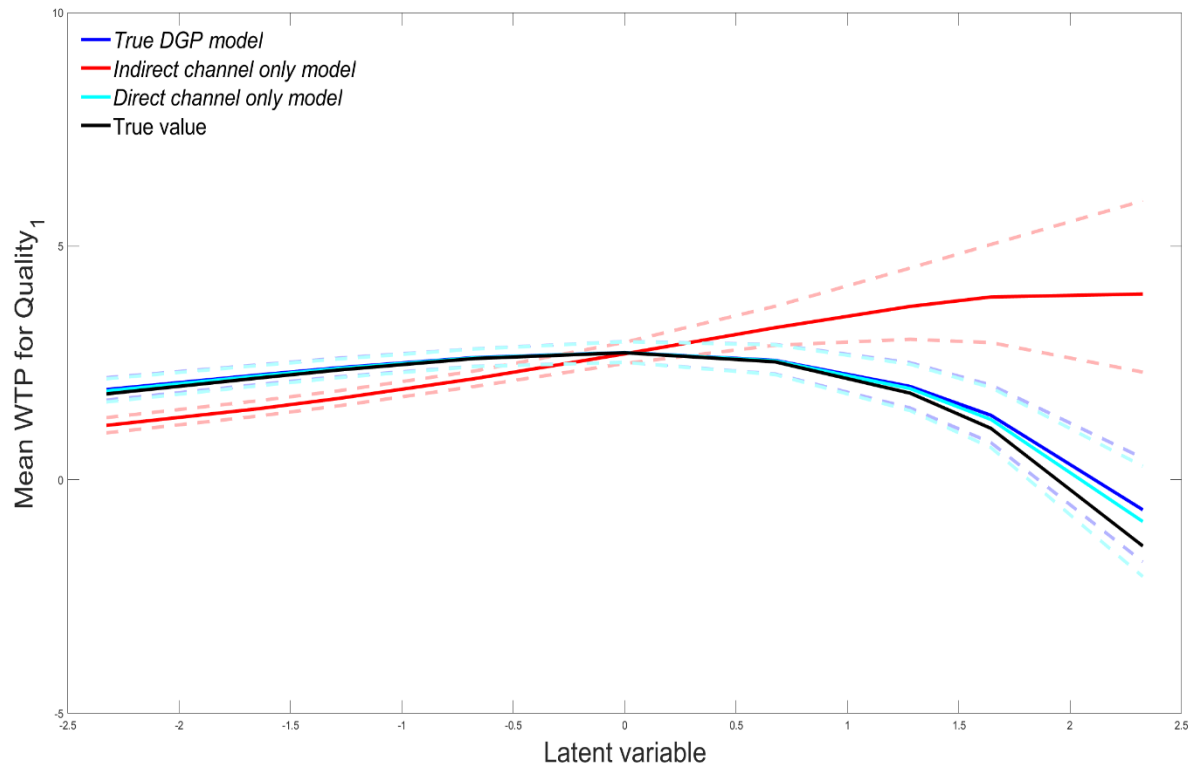
Results – observed heterogeneity

- Relation between WTP and socio-demographic variable is recovered well with reduced form model (MXL)



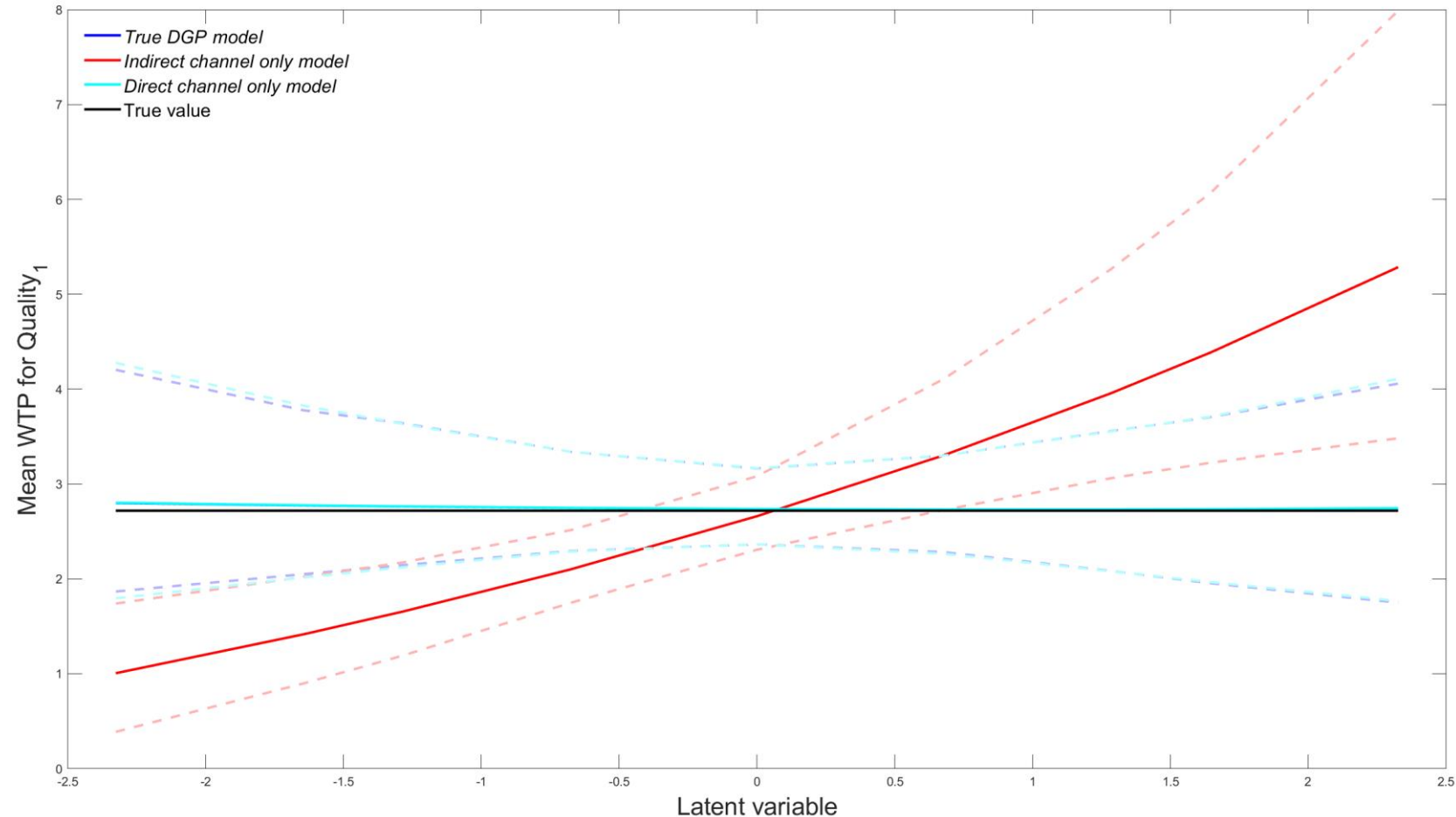
Results – observed heterogeneity

- Relation between WTP and latent variable is also biased, but only for '*Indirect channel only*' model



Results – observed heterogeneity (LV not significant)

- For '*Indirect channel only*' model we found spurious significance of LV



Conclusions

- We find that misspecifying unobserved heterogeneity or observed heterogeneity biases inference in hybrid choice models
 - In the former case, mean/median WTP is largely affected as well as relationship between LV and WTP
 - In the latter case, mostly relationship between LV and WTP is affected
 - It can lead to spurious significance of LV
 - Also affects the results for socio-demographic variables
- Reduced form model can be use to guide choice of the specification
- Future work
 - Empirical example on real dataset