

Correcting Selection Bias in Non-probability Two-Phase Payment Survey

Heng Chen¹, John Tsang²

¹Bank of Canada, ²University of Ottawa

The 2020 Cash Alternative Survey Wave 2 (CASW2)

What is CASW2?

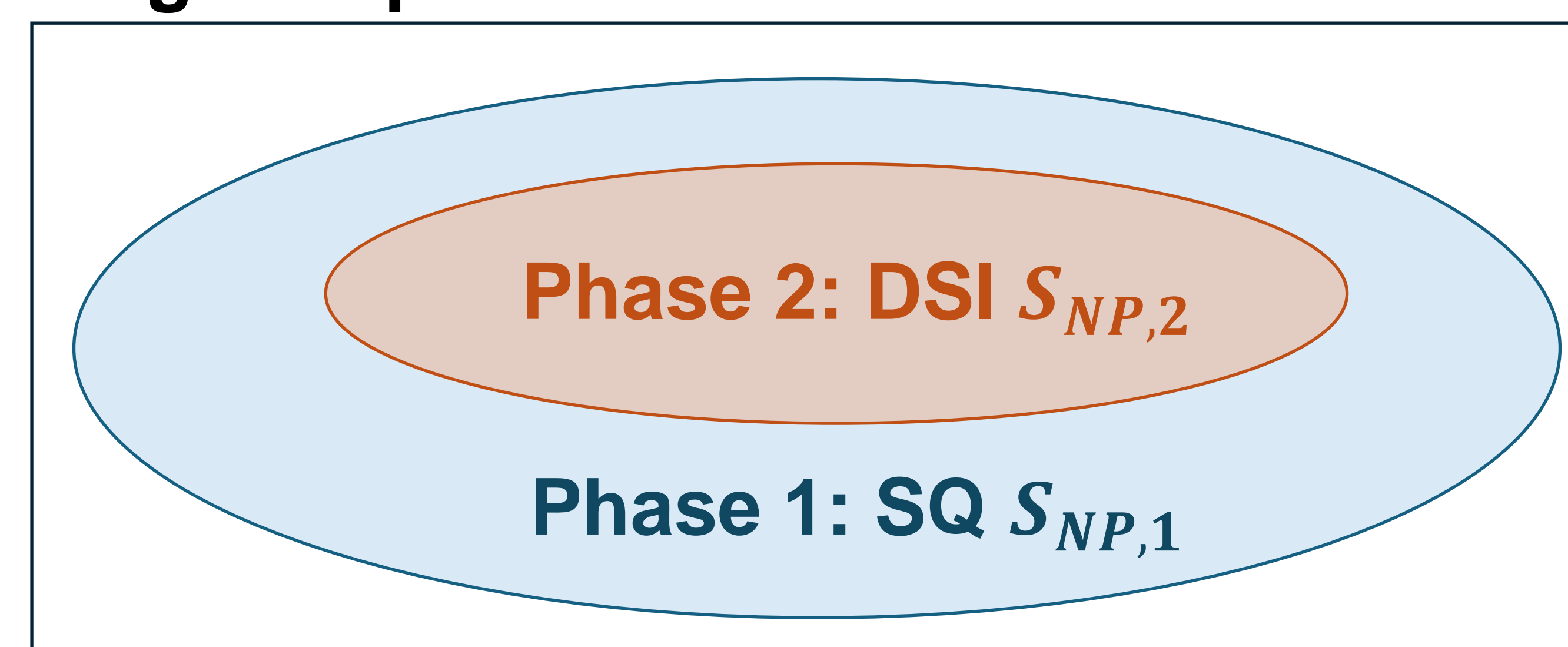
CASW2 collects information about cash holdings and Canadians' daily use of cash.

Two Phases in CASW2

- Phase 1: Survey Questionnaire (SQ)
- Phase 2: Diary Survey Instrument (DSI)

Figure 1. Every DSI-respondent answered the SQ

Target Population: Canadian Adults



Data Structures for Phase 1

Table 1. Augment the SQ with Prob. Sample CPSS 5

Reference Dataset	Target Dataset
Prob. Sample (CPSS 5 from StatsCan)	Phase-1 Sample (SQ)
S_P	S_{NP}
Auxiliary variables x_1	✓
Study variables y_1	✗
Phase-1 weight w_1	✓

Data Structures for Phase 2

Table 2. Augment the DSI with the SQ

	Reference Datasets		Target Dataset
	Probability Sample	Phase-1 SQ	Phase-2 DSI
	S_P	S_{NP}	S_{NP} with $I_2 = 1$
Auxiliary variables z	✗	✓	✓
Auxiliary variables x_1	✓	✓	✓
Phase-1 study variables y_1	✗	✓	✓
Phase-2 study variables y_2	✗	✗	✓
Phase-1 weight w_1	✓	(Estimated)	(Estimated)
Phase-2 weight w_2	✗	-	✗

Main Assumptions

- A1:** Non-informative selection for both phases
- A2:** No hard-core non-participants
- A3:** Independent sample selection
- A4:** Logistic models for selection probabilities

Overview

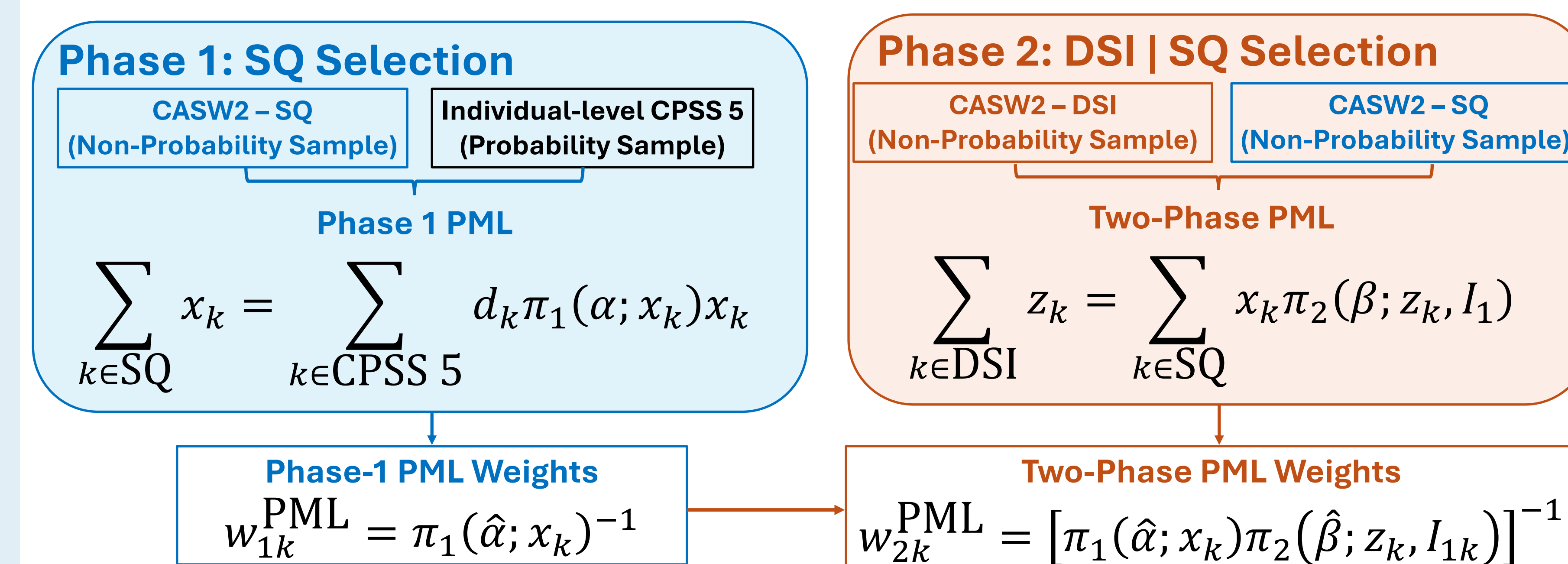
- Extend the pseudo-maximum-likelihood (PML) weight estimation from Chen, Li and Wu (2020) to two-phases.
- Propose a pseudo-population bootstrap (PPB) procedure for variance estimation
- Application: The Bank of Canada 2020 CASW2

Contributions

- This paper is one of the first for **point and variance estimation in Phase 2** under data integration with non-probability and probability samples.
- We are one of the first to use the **pseudo-population bootstrap (PPB)** to estimate variance for non-probability survey samples.

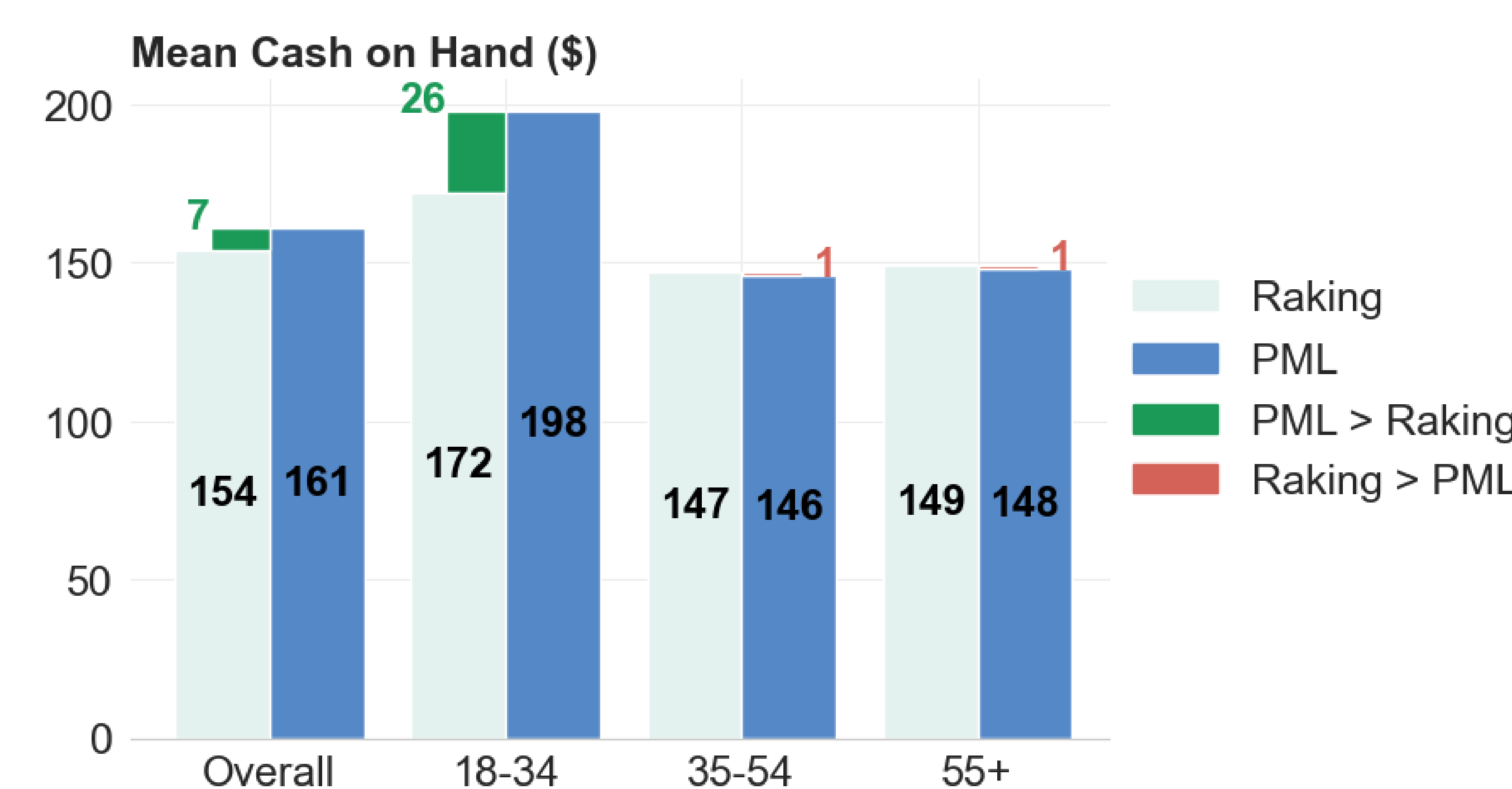
Weight Estimation Procedure

Figure 2. Weights Parallel to Two-Phase Prob. Design



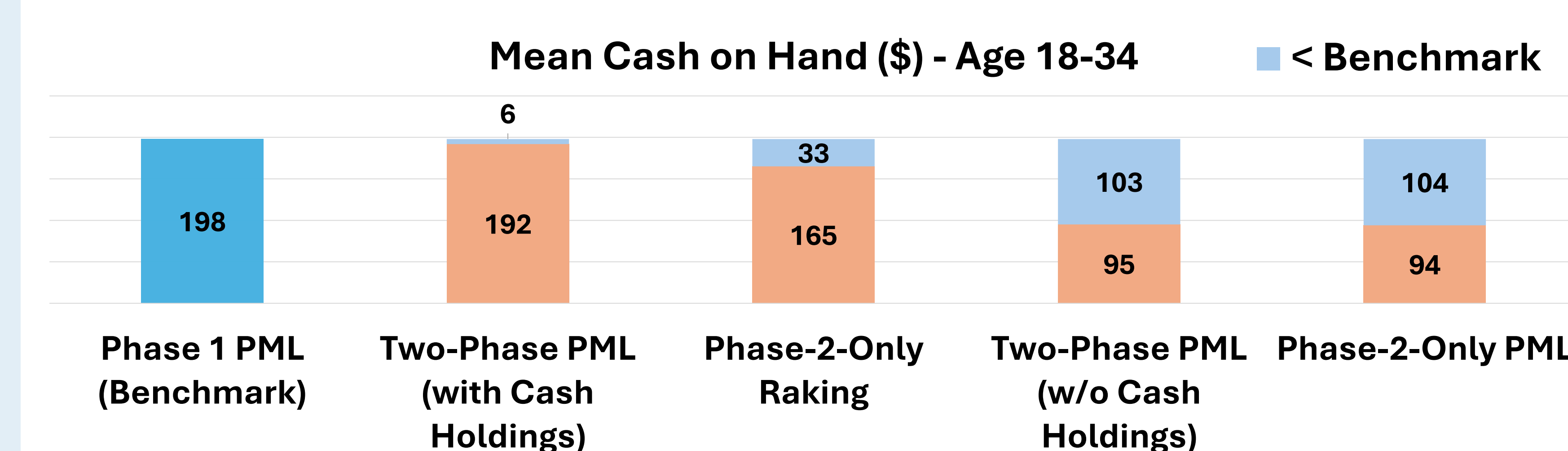
Point Estimation in Phases 1 and 2

Figure 3. PML Addresses Bias from the Low-Response Age 18-34 Better than Raking



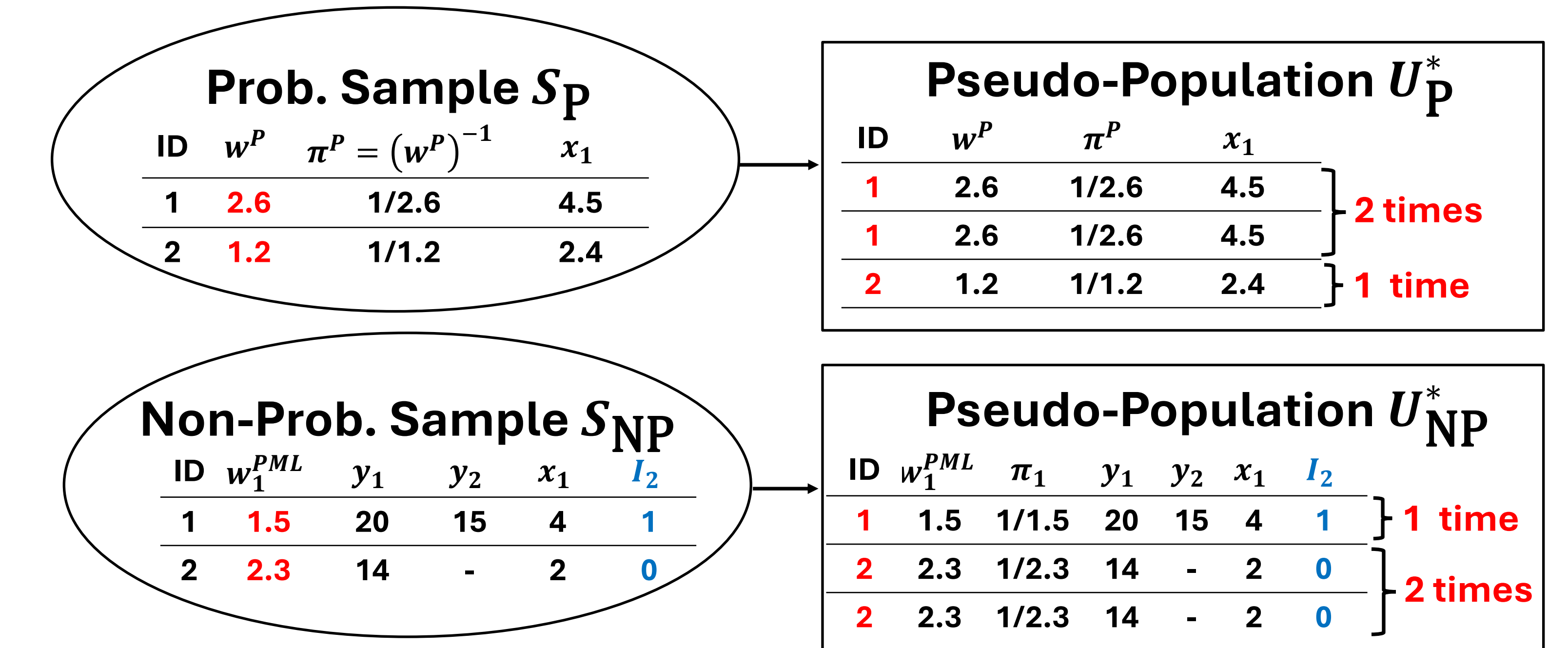
Validated Phase-1 PML weights by comparing estimates from SQ and CPSS 5 at the overall and the age-group level.

Figure 4. Cash Holdings from Phase 1 are Valuable for Phase-2 Estimation

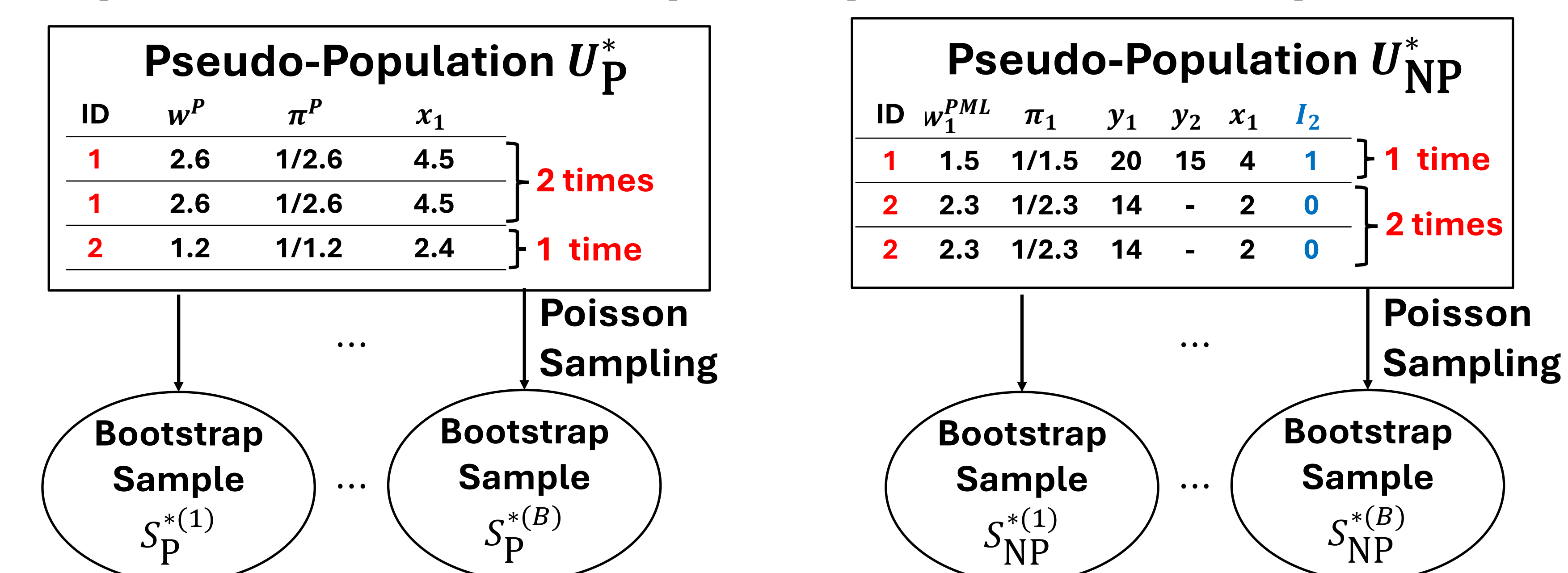


PPB Procedure for Variance Estimation

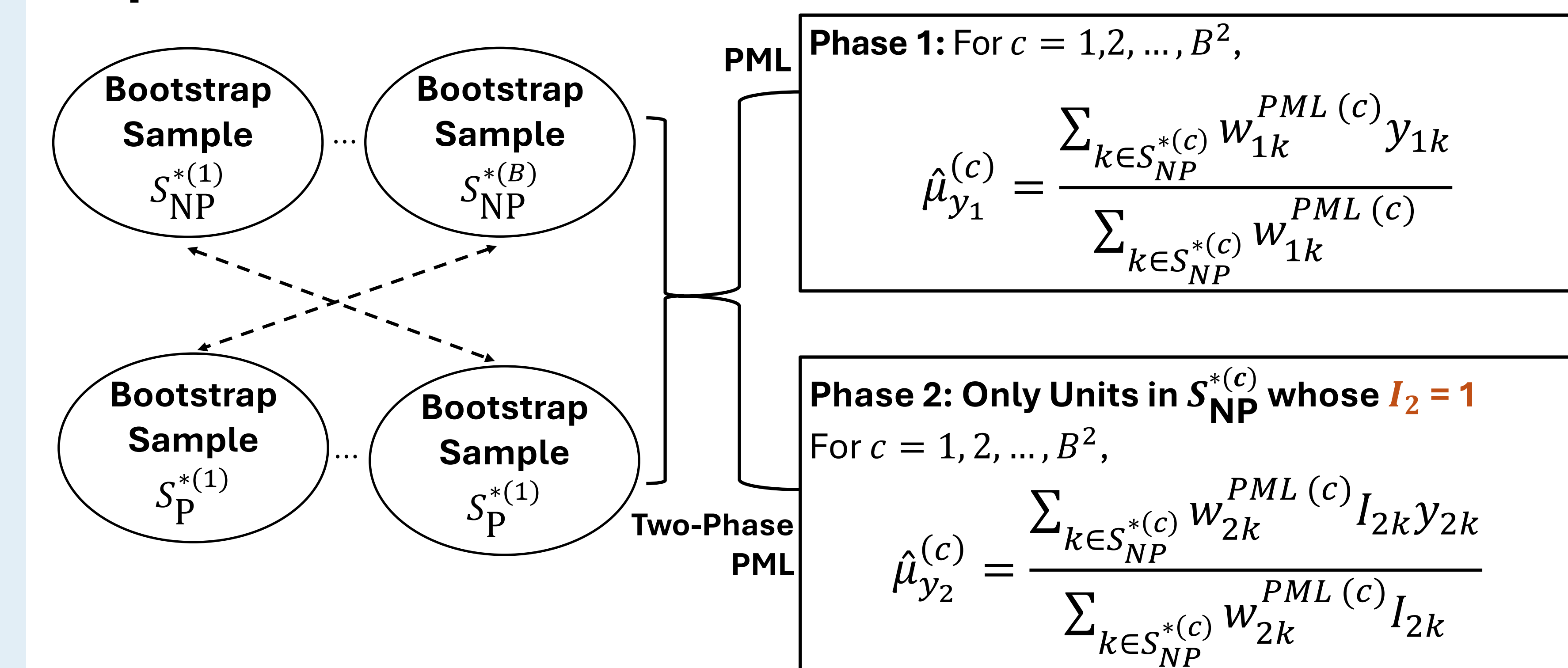
Step 1: Create Pseudo-Populations from S_P and $S_{NP,1}$



Step 2: Create Bootstrap Samples Pseudo-Populations



Step 3: Variance Estimation - Phase-1&Two-Phase PML



- Use the customary variance formula to estimate $\text{Var}[\hat{\mu}_{y_1}]$ and $\text{Var}[\hat{\mu}_{y_2}]$.
- Follow the same principle as the reverse approach (Shao and Steel, 1999 and Haziza and Vallée, 2020) for variance estimation with non-response to approximate variance for Phase 2 under two-phase PML.
 - A5: Negligible phase-1 sampling fraction**
 - A6: Strong invariance property** (Beaumont and Haziza, 2016): Phase-2 selection probabilities are determined by fixed respondent characteristics.

References

- Beaumont, J. F., and Haziza, D. (2016). A note on the concept of invariance in two-phase sampling designs. *Survey Methodology*, 42(2), 319-323.
- Chen, Y., Li, P., and Wu, C. (2020). Doubly robust inference with non-probability survey samples. *Journal of the American Statistical Association*, 115, 2011-2021.
- Haziza, D., and Vallée, A. A. (2020). Variance estimation procedures in the presence of singly imputed survey data: a critical review. *Japanese Journal of Statistics and Data Science*, 3, 583-623.
- Shao, J., and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94(445), 254-265.

This work is financially supported by the Bank of Canada and Mitacs through the Mitacs Accelerate program. The views expressed in this presentation are those of the authors and do not necessarily represent the official views of the Bank of Canada. All remaining errors are solely the responsibility of the authors.