Online Supplement

Simulation Study S1

A simple simulation design is carried out to evaluate the performance of the non-parametric classification (NPC, Chiu & Douglas, 2013) method with that of the commonly used profile estimation method MAP when the posterior weighted Kullback–Leibler Information Index (PWKL, Cheng, 2009) is used as the item selection method in CD-CAT.

Design

Item bank generation. We consider the DINA, DINO and RRUM model, respectively, with the number of attributes $K \in \{3, 4\}$ and the item bank size J =350.

The fixed test length L = 30. In the study, in order to investigate the influence of Q-matrix on the NPC method, we will consider two kinds of Q-matrix structures. Like Wang (2013), the one Q-matrix followed a simple structure, in which one Kth of the items exclusively measured each of the K attributes. The other type of Q-matrix followed complex structure. For complex Q-matrices, every entry was accompanied by a random number from Uniform(0,1). If the random number was smaller than 0.5, then the corresponding Q-matrix entry was 1, indicating that the item required the attribute. The corresponding Q-matrix entry was set to be 0, otherwise. It was noted that every item was constrained to measure at least one of the five attributes in order to avoid trivial rows in the Q-matrix.

Another critical factor affecting the classification results is the distribution of item parameters. From a practical point of view, it is important to investigate conditions in which we can obtain good non-parametric classifications. Therefore, the guessing and slipping parameters in the simulations were generated from uniform distribution U(0, *Max.s*), where *Max.s* was set to be 0.1, 0.3, or 0.5, denoting low, medium, and high perturbations.

Examinees generation. Like Chiu and Douglas (2013), $N = 1000$ examinees' attribute patterns were generated in two different ways. The first sampled attribute patterns, $\boldsymbol{\alpha}$, are from a uniform distribution on 2^K possible values, each with the probability $1/2^K$. The second method, as known as multivariate normal threshold model, was used to mimic a realistic situation where attributes were correlated and of unequal prevalence. The discrete $\boldsymbol{\alpha}$ were linked to an underlying multivariate normal distribution, $\boldsymbol{\theta}_\mathbf{i} \sim MVN(\mathbf{0}_K, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$, had the structure as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

and ρ was set to be 0.5. Let $\boldsymbol{\theta}_\mathbf{i} = (\theta_{i1}, \theta_{i2}, ..., \theta_{iK})'$ denote the K-dimensional vector of latent continuous scores for examinee i. The attribute pattern $\boldsymbol{\alpha}_\mathbf{i} = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{iK})'$ was determined by

$$\alpha_{ik} = \begin{cases} 1, & \theta_{ik} \geq \Phi^{-1}\left(\dfrac{k}{K+1}\right); \\ 0, & \text{otherwise}. \end{cases} \tag{S4}$$

Item selection algorithms. The PWKL was used as the item selection method.

Parameter estimation. The attribute pattern estimates, $\hat{\boldsymbol{\alpha}}$, are obtained via NPC method and maximum a posteriori (MAP) method with the uniform prior (i.e., $U(0,1)$).

Stopping rule. The fixed-length method ($L = 30$) was used to terminate the algorithms.

Therefore, we had 2 (Q-matrix structure) × 2 (number of attributes) × 3 (data generation models) × 3 (bank information) × 2 (attribute structure) = 72 data generation conditions for the simulation study. For each condition, 30 replications were generated.

Results

The tables below report not only *PARs* and *AARs*, but also their 'relative efficiency' for the

NPC versus MAP method. Note that the index's "relative efficiency" is defined as the ratio of the NPC versus MAP indices, and represents the proportion of individual attributes that were classified correctly.

Table S1 presents the *AAR*s, *PAR*s, and their relative efficiencies for the NPC method and MLE method when the data conformed to the DINA model. As noted by Chiu and Douglas (2013), "*In the case of the DINA model, the ideal response pattern will always be the most like pattern, unless slipping and guessing values exceed 0.5.*" Both methods produce nearly perfect classifications when slipping and guessing values are less than 0.3. All relative efficiencies have values less than 1, indicating that NPC method outperformed the MAP method in almost all conditions. From the *PAR*s, we can see the NPC method classified at most 14.95% more examinees into the correct proficiency classes than the parametric method. The table also indicates that larger numbers of attributes, simple *Q* matrix, and poor item quality each caused the mean *PAR*s and *AAR*s to decrease. Meanwhile, when attribute patterns conformed to the multivariate normal threshold model, the mean *PAR*s and *AAR*s increased. Because the multivariate normal threshold model incorporates a far more realistic scenario than the uniform distribution model (Chiu & Douglas, 2013), these findings suggest that when the data conform to the DINA model, the NPC method appears to be the best choice.

Table S2 summarizes the results for the data generated from the DINO model. It contains same patterns as noted in Table S1. When the data conformed to the DINA model, the NPC and MAP methods performed well when slipping and guessing values are less than 0.3. When slipping and guessing values increasing, *PAR*s and *AAR*s of both approaches tend to decline. We notice that the performance of the two classification methods also depends on the size of *K*. The results showed that larger number of attributes resulted in smaller *PAR* and *AAR* scores for both methods.

For the impressive performance of the NPC method when the data conformed to the DINA and DINO model, a heuristic explanation can be derived from the proof given by Chiu and Douglas (2013): "*as long as the slipping and guessing parameters do not exceed 0.5, the ideal response pattern will be the most like choice for the proficiency class*".

From the effectiveness of the NPC method and the MAP method described, when the data were generated from the RRUM model in Table S3, *PAR* and *AAR* values under the two classification methods display the same trends as found in Tables S1 and S2. However, some relative efficiency values in Table S3 are approximately equal to 1, indicating that the NPC method sometimes performs about as well as MAP method. The NPC method appears generally less tolerant of larger slipping and guessing parameters in the RRUM model, which maybe because of the multiplicative effect of the larger slipping and guessing operating at the subtask level (Chiu & Douglas, 2013).

Table S1. Agreement of classification between the NPC method and MAP method with data generated from the DINA model in CD-CAT.

Q-structure	K	*Max.s*	MAP		NPC		Relative Efficiency (MAP/NPC)	
			PAR	*AAR*	*PAR*	*AAR*	*PAR*	*AAR*
			Uniform Attribute Patterns					
Simple	3	0.1	0.9973	0.9985	0.9980	0.9989	0.9993	0.9996
		0.3	0.8917	0.9620	0.8947	0.9696	0.9966	0.9922
		0.5	0.8215	0.9010	0.9398	0.9647	0.8741	0.9340
	4	0.1	0.8631	0.9650	0.8633	0.9657	0.9998	0.9993
		0.3	0.7775	0.8618	0.9270	0.9467	0.8387	0.9103
		0.5	0.7245	0.8431	0.8086	0.8882	0.8960	0.9492
Complex	3	0.1	0.9974	0.9991	0.9984	0.9993	0.9990	0.9998
		0.3	0.9811	0.9909	0.9914	0.9960	0.9895	0.9949
		0.5	0.9618	0.9842	0.9883	0.9958	0.9732	0.9884
	4	0.1	0.9290	0.9581	0.9870	0.9926	0.9412	0.9653
		0.3	0.9181	0.9608	0.9628	0.9828	0.9535	0.9776
		0.5	0.8485	0.9308	0.9050	0.9593	0.9376	0.9703
			Multivariate Normal Attribute Patterns					
Simple	3	0.1	0.9732	0.9910	0.9742	0.9913	0.9990	0.9997
		0.3	0.9288	0.9752	0.9342	0.9767	0.9943	0.9985
		0.5	0.8841	0.9230	0.8889	0.9275	0.9945	0.9951
	4	0.1	0.9527	0.9876	0.9588	0.9885	0.9937	0.9990
		0.3	0.8666	0.9584	0.8684	0.9590	0.9979	0.9994
		0.5	0.7821	0.8498	0.9222	0.9421	0.8481	0.9020
Complex	3	0.1	0.9954	0.9964	0.9992	0.9995	0.9962	0.9969
		0.3	0.9852	0.9938	0.9937	0.9970	0.9914	0.9968
		0.5	0.9597	0.9864	0.9778	0.9919	0.9814	0.9945
	4	0.1	0.9945	0.9986	0.9969	0.9991	0.9976	0.9995
		0.3	0.9777	0.9885	0.9806	0.9899	0.9970	0.9986
		0.5	0.9277	0.9678	0.9448	0.9723	0.9820	0.9954

Note. …..

Table S2. Agreement of classification between the NPC method and MAP method with data generated from the DINO model in CD-CAT.

Q-structure	K	*Max.s*	MAP		NPC		Relative Efficiency (MAP/NPC)	
			PAR	*AAR*	*PAR*	*AAR*	*PAR*	*AAR*
Uniform Attribute Patterns								
Simple	3	0.1	0.9799	0.9932	0.9809	0.9957	0.9990	0.9975
		0.3	0.9321	0.9762	0.9572	0.9780	0.9737	0.9982
		0.5	0.7911	0.9270	0.7925	0.9311	0.9982	0.9956
	4	0.1	0.9559	0.9884	0.9592	0.9895	0.9966	0.9989
		0.3	0.8833	0.9669	0.8912	0.9733	0.9911	0.9934
		0.5	0.7855	0.9136	0.7859	0.9180	0.9995	0.9952
Complex	3	0.1	0.9964	0.9987	0.9975	0.9994	0.9989	0.9993
		0.3	0.9396	0.9769	0.9516	0.9776	0.9874	0.9993
		0.5	0.8884	0.9548	0.8944	0.9560	0.9933	0.9987
	4	0.1	0.9757	0.9885	0.9765	0.9895	0.9991	0.9990
		0.3	0.9064	0.9747	0.9284	0.9798	0.9764	0.9948
		0.5	0.8677	0.9521	0.8759	0.9558	0.9906	0.9961
Multivariate Normal Attribute Patterns								
Simple	3	0.1	0.9712	0.9934	0.9725	0.9959	0.9987	0.9975
		0.3	0.9315	0.9707	0.9525	0.9799	0.9779	0.9906
		0.5	0.8760	0.9510	0.8827	0.9562	0.9924	0.9946
	4	0.1	0.9447	0.9852	0.9470	0.9864	0.9976	0.9988
		0.3	0.9156	0.9722	0.9213	0.9784	0.9938	0.9937
		0.5	0.8410	0.9431	0.8485	0.9481	0.9912	0.9947
Complex	3	0.1	0.9957	0.9918	0.9965	0.9925	0.9993	0.9993
		0.3	0.9491	0.9826	0.9499	0.9882	0.9992	0.9943
		0.5	0.8920	0.9604	0.9151	0.9656	0.9748	0.9946
	4	0.1	0.9603	0.9816	0.9636	0.9857	0.9966	0.9958
		0.3	0.9210	0.9786	0.9592	0.9828	0.9602	0.9957
		0.5	0.8403	0.9581	0.8521	0.9598	0.9862	0.9982

Table S3. Agreement of classification between the NPC method and MAP method with data generated from the RRUM model in CD-CAT.

Q-structure	K	*Max.s*	MAP		NPC		Relative Efficiency (MAP/NPC)	
			PAR	AAR	PAR	AAR	PAR	AAR
Uniform Attribute Patterns								
Simple	3	0.1	0.9391	0.9796	0.9395	0.9797	0.9996	0.9999
		0.3	0.8141	0.9372	0.8146	0.9367	0.9994	1.0005
		0.5	0.7541	0.9135	0.7547	0.9132	0.9992	1.0004
	4	0.1	0.8712	0.9671	0.8708	0.9664	1.0005	1.0007
		0.3	0.7509	0.9114	0.7556	0.9165	0.9938	0.9944
		0.5	0.6442	0.8663	0.6439	0.8661	1.0005	1.0003
Complex	3	0.1	0.9940	0.9911	0.9934	0.9901	1.0006	1.0010
		0.3	0.9486	0.9739	0.9480	0.9735	1.0006	1.0004
		0.5	0.8693	0.9492	0.8686	0.9485	1.0008	1.0007
	4	0.1	0.9682	0.9921	0.9706	0.9933	0.9976	0.9988
		0.3	0.7881	0.9408	0.7889	0.9431	0.9990	0.9976
		0.5	0.7490	0.9131	0.7493	0.9134	0.9996	0.9997
Multivariate Normal Attribute Patterns								
Simple	3	0.1	0.9281	0.9759	0.9287	0.9766	0.9993	0.9993
		0.3	0.7341	0.9083	0.7344	0.9080	0.9995	1.0003
		0.5	0.6653	0.8774	0.6654	0.8768	0.9997	1.0007
	4	0.1	0.8506	0.9616	0.8505	0.9614	1.0001	1.0002
		0.3	0.6363	0.9000	0.6392	0.9007	0.9954	0.9993
		0.5	0.5883	0.8640	0.5893	0.8664	0.9982	0.9972
Complex	3	0.1	0.9980	0.9990	0.9979	0.9988	1.0001	1.0002
		0.3	0.9099	0.9610	0.9094	0.9627	1.0006	0.9982
		0.5	0.7638	0.9134	0.7653	0.9138	0.9981	0.9995
	4	0.1	0.9640	0.9958	0.9681	0.9967	0.9957	0.9991
		0.3	0.8341	0.9579	0.8342	0.9581	0.9998	0.9998
		0.5	0.7019	0.9142	0.7090	0.9155	0.9899	0.9986

References

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing. *Psychometrika*, *74*, 619-632.

Chiu, C. Y., & Douglas, J. A. (2013). A non-parametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, *30*, 225-250.

Ma, W., & de la Torre, J. (2017). GDINA: The generalized DINA model framework. *R package version 1.4.2*. Retrieved from https://CRAN.R-project.org/package=GDINA.

Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Journal of Educational Measurement*, *73*(6), 1017-1035.

Zheng, Y., & Chiu, C. (2015). NPCD: An R package for non-parametric methods for cognitive diagnosis. *R package version 1.0-9*. Retrieved from http://CRAN.R-project.org/package=NPCD

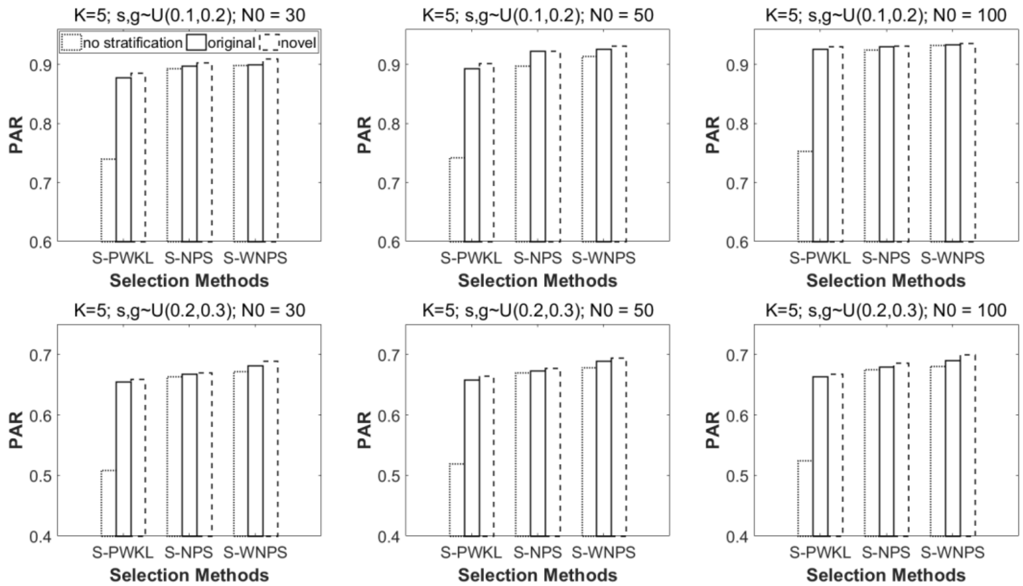

Figure S1. *PARs* under DINA model when $K = 5$ and number of strata = 3

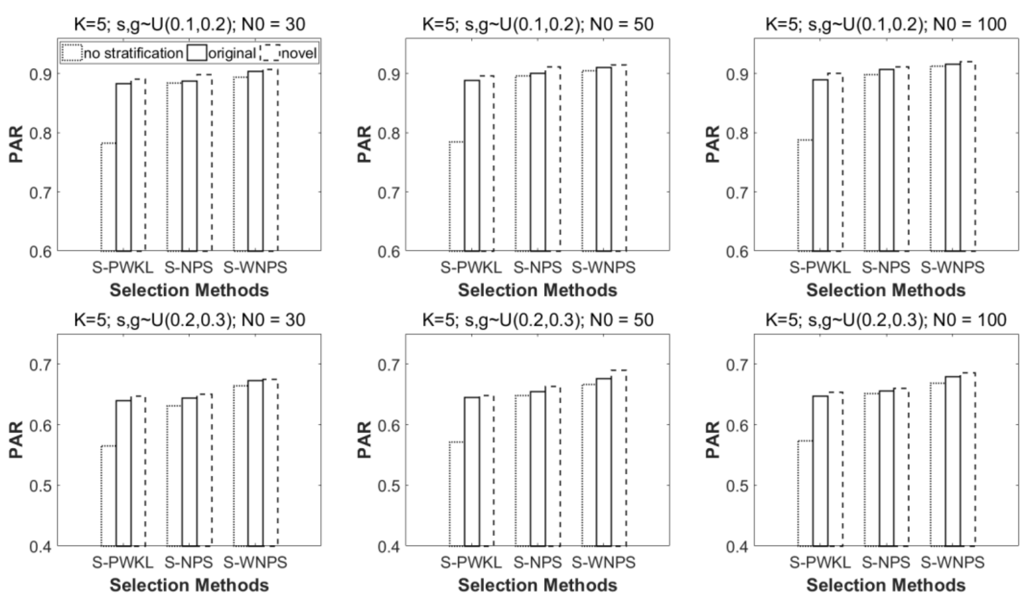

Figure S2. *PARs* under DINO model when $K = 5$ and number of strata = 3

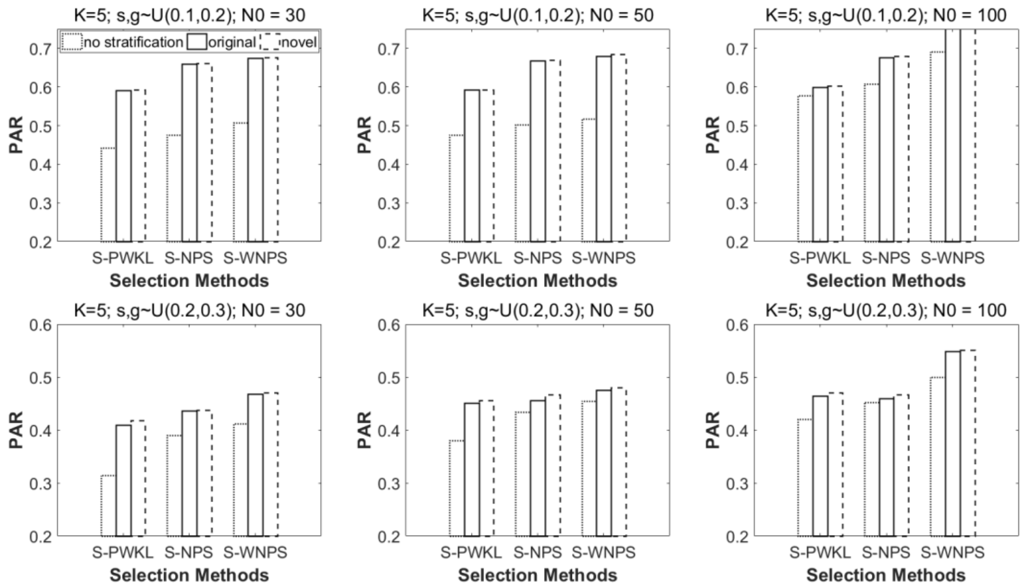

Figure S3. *PAR*s under the RRUM model when K = 5 and number of strata = 3

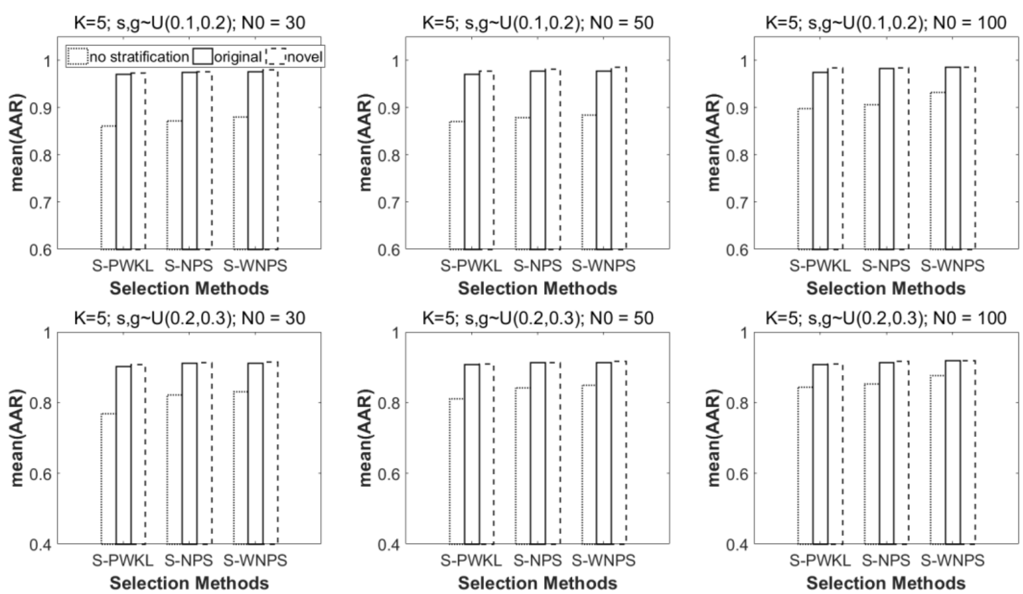

Figure S4. mean of *AAR*s under the DINA model when K = 5 and number of strata = 3

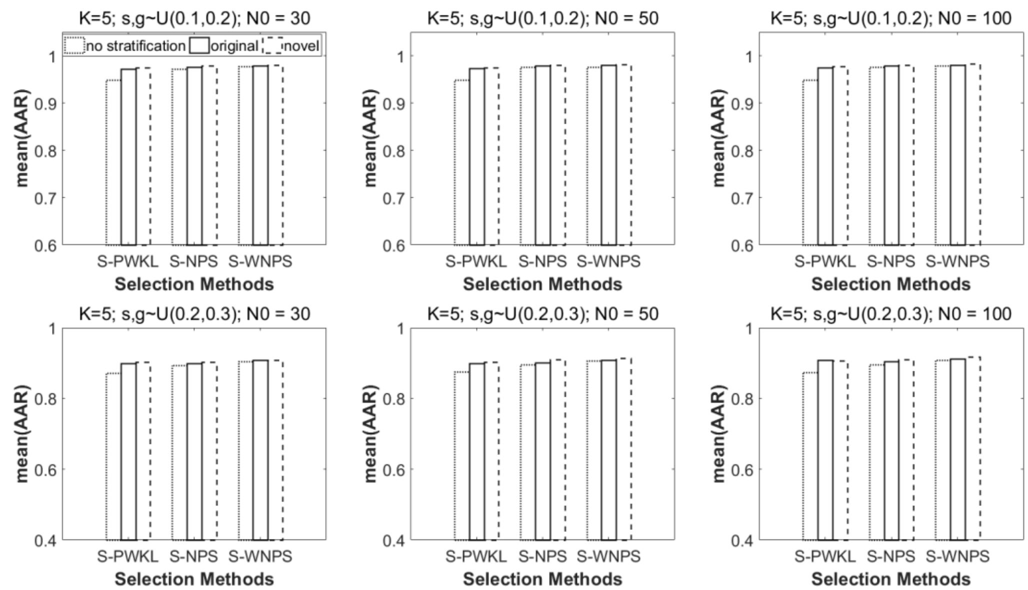

Figure S5. mean of *AAR*s under the DINO model when K = 5 and number of strata = 3

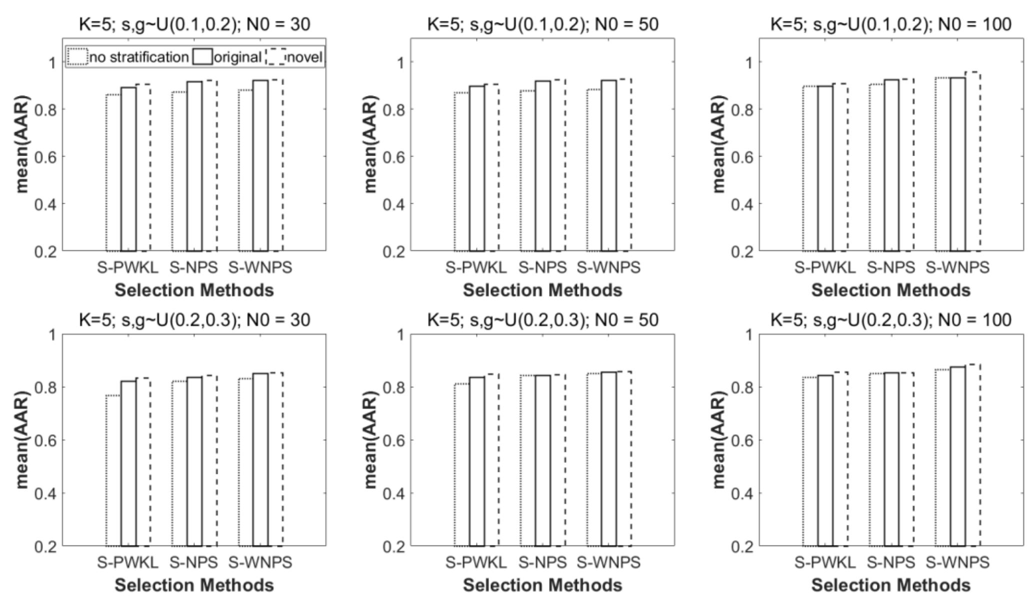

Figure S6. mean of *AAR*s under the RRUM model when $K = 5$ and number of strata = 3

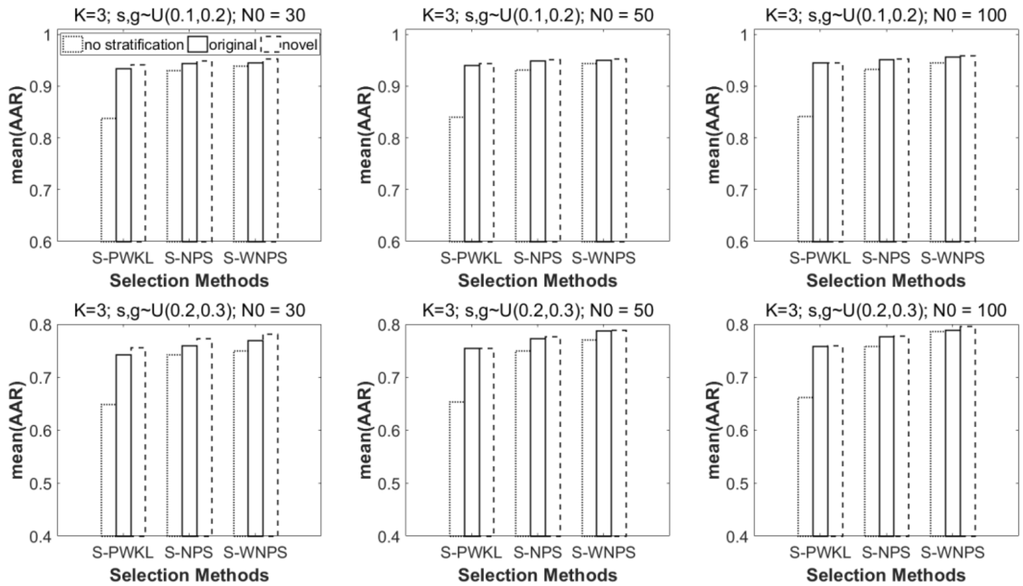

Figure S7. *PARs* under DINA model when $K = 3$ and number of strata = 5

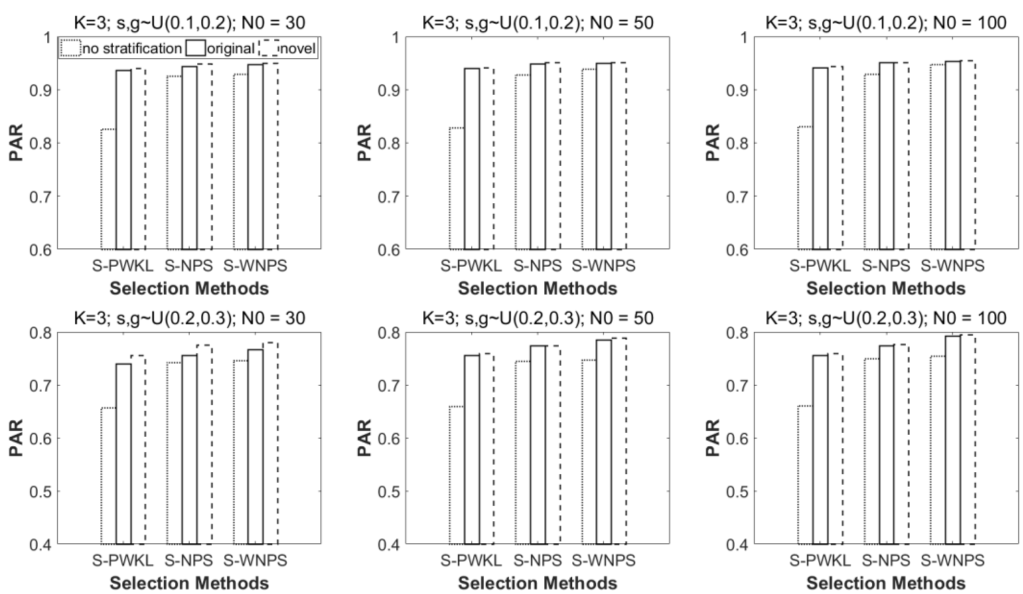

Figure S8. *PARs* under DINO model when $K = 3$ and number of strata = 5

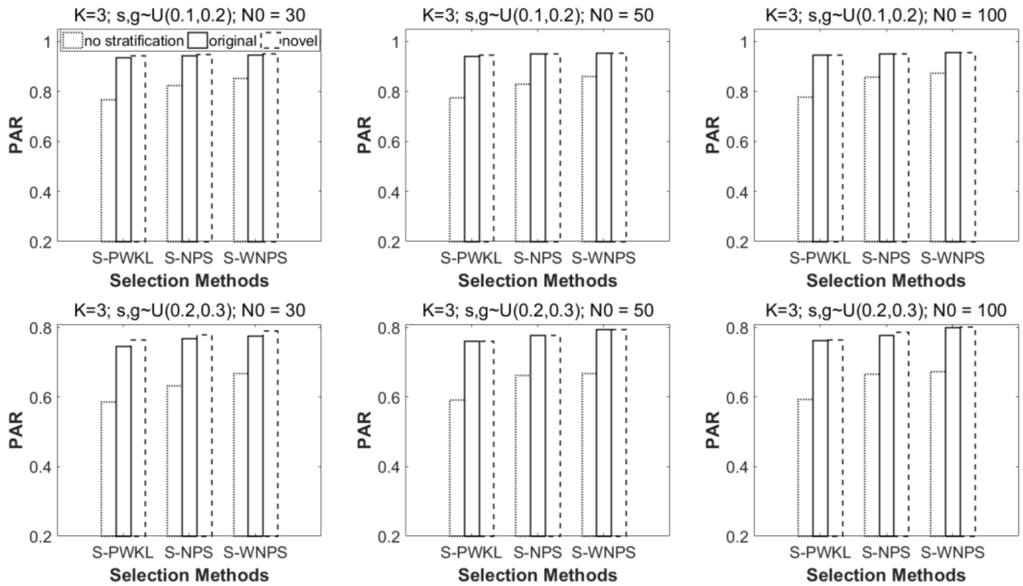

Figure S9. *PAR*s under the RRUM model when K = 3 and number of strata = 5

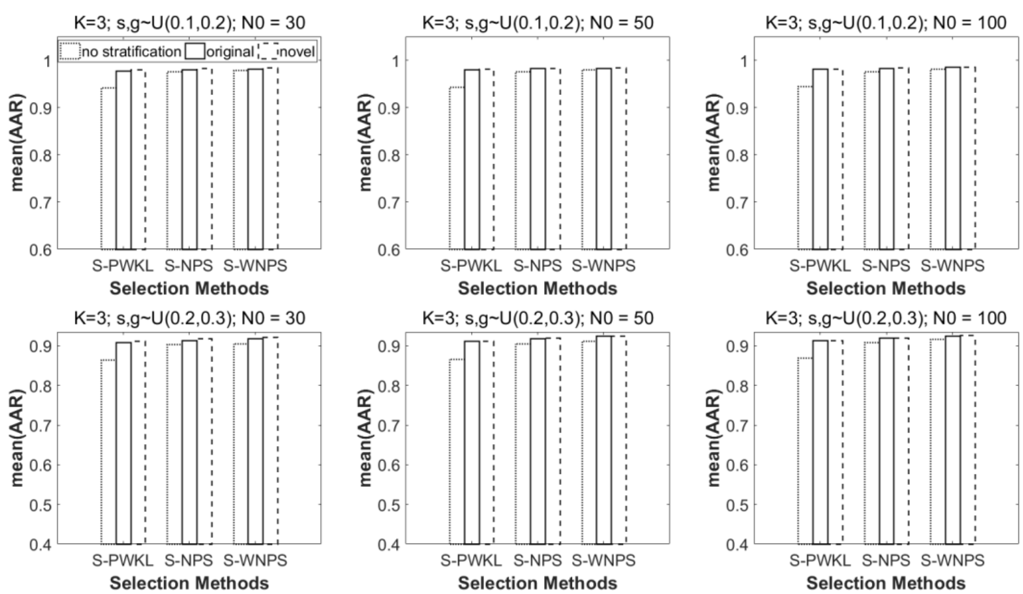

Figure S10. mean of *AAR*s under the DINA model when K = 3 and number of strata = 5

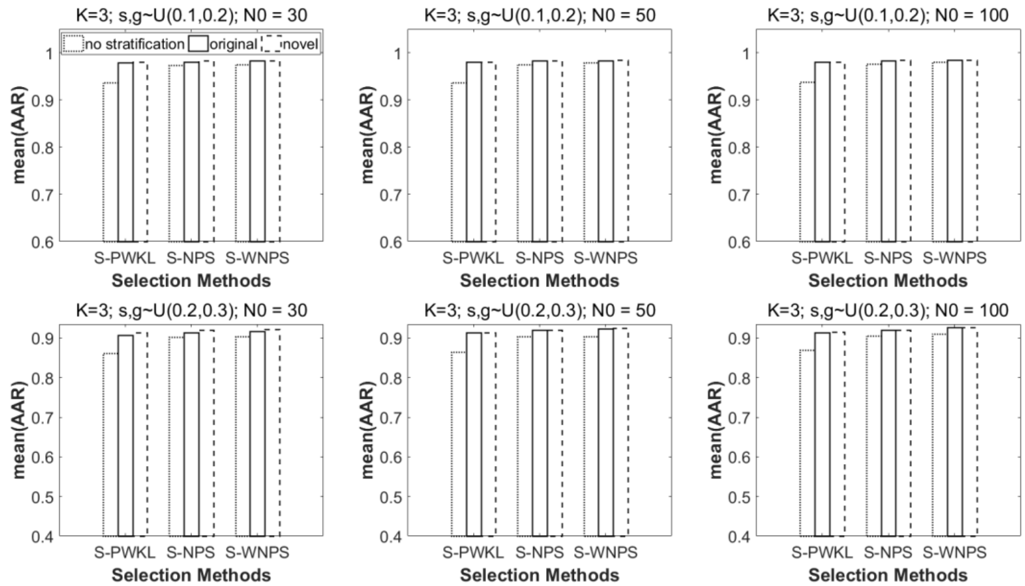

Figure S11. mean of *AAR*s under the DINO model when K = 3 and number of strata = 5

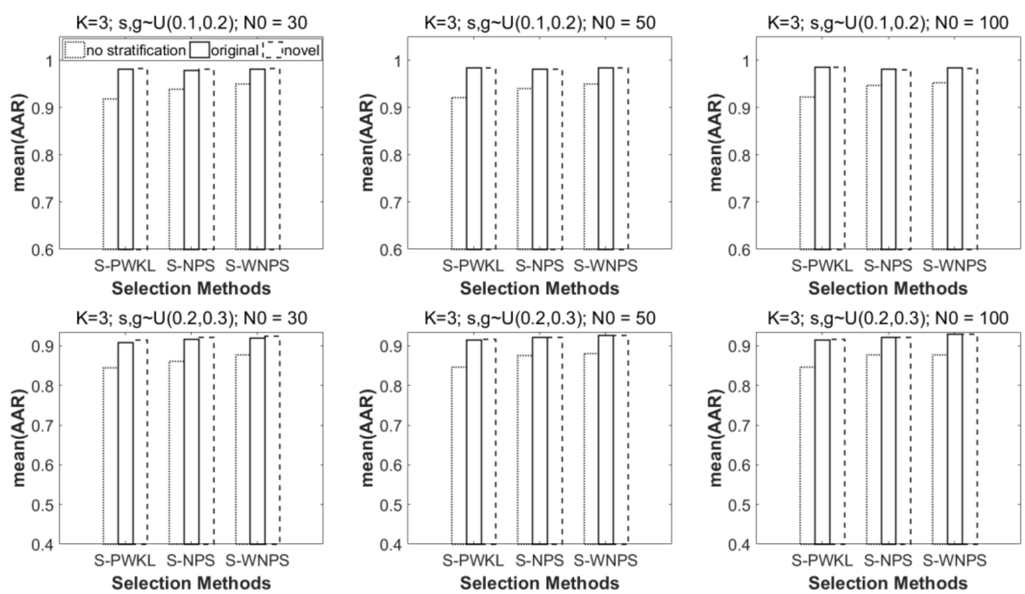

Figure S12. mean of *AAR*s under the RRUM model when K = 3 and number of strata = 5

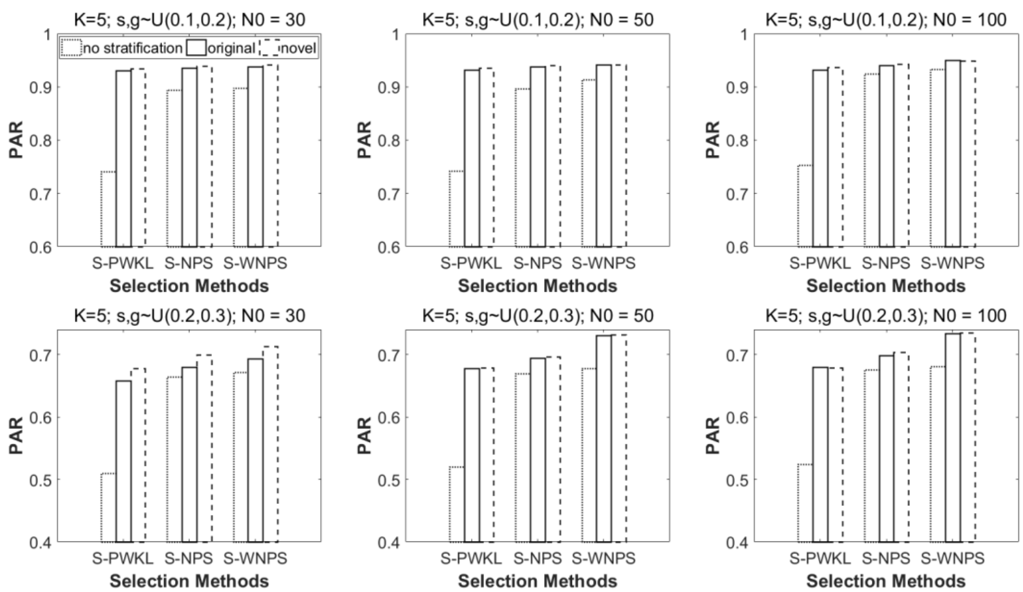

Figure S13. *PARs* under DINA model when K = 5 and number of strata = 5

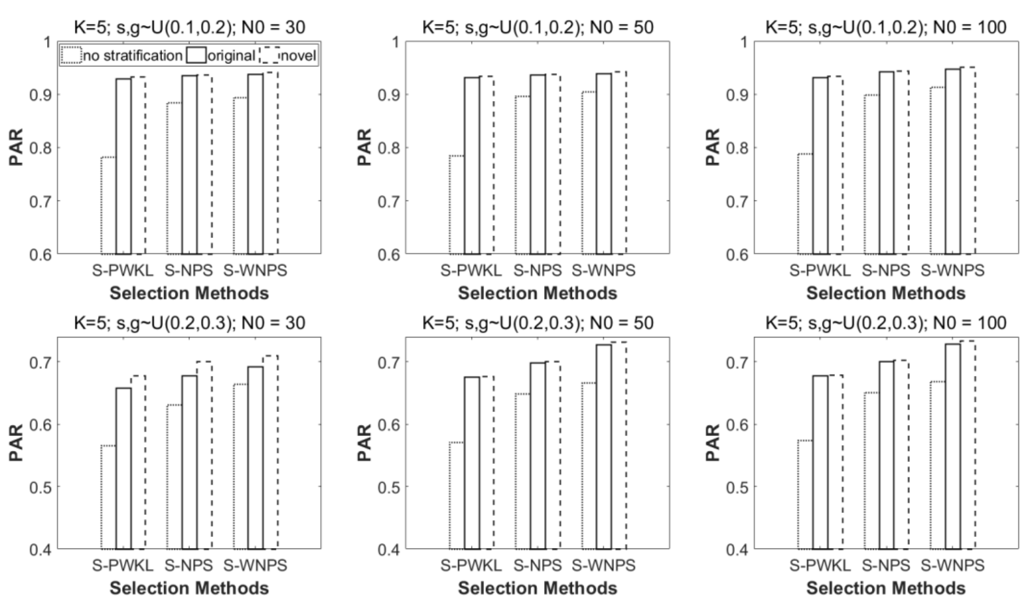

Figure S14. *PARs* under DINO model when *K* = 5 and number of strata = 5

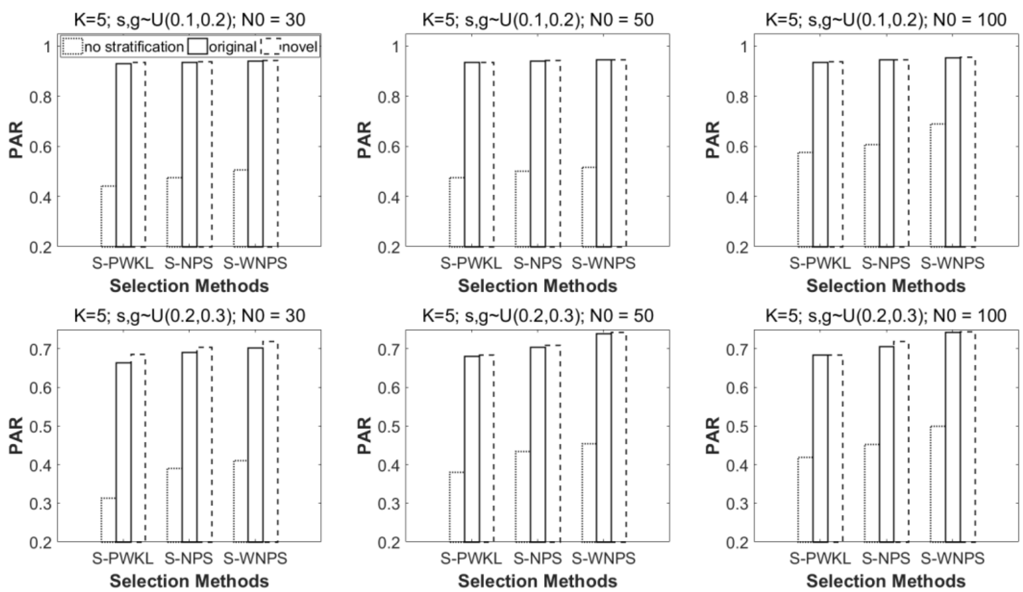

Figure S15. *PAR*s under the RRUM model when *K* = 5 and number of strata = 5

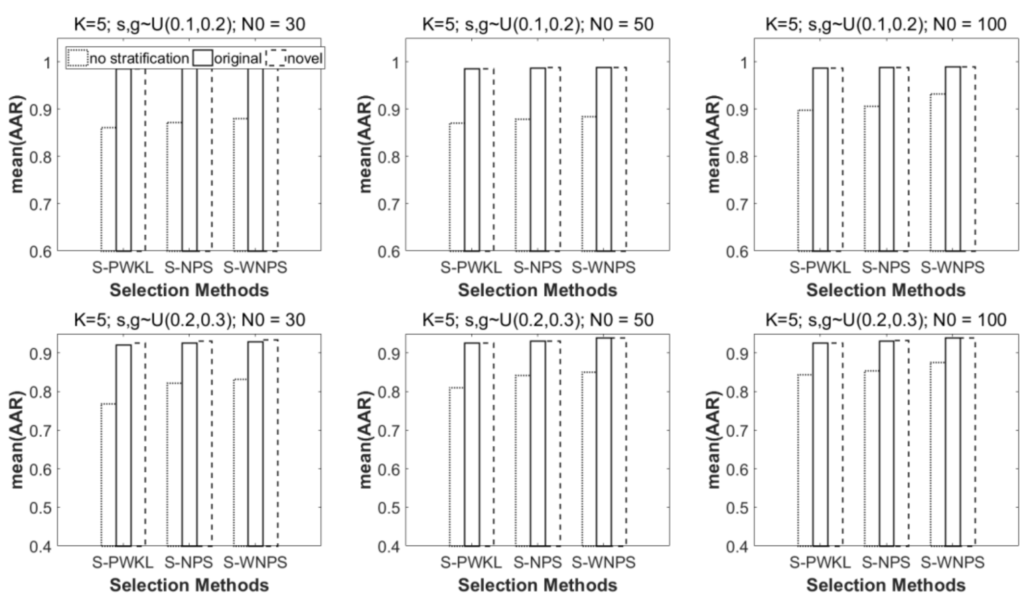

Figure S16. mean of *AAR*s under the DINA model when K = 5 and number of strata = 5

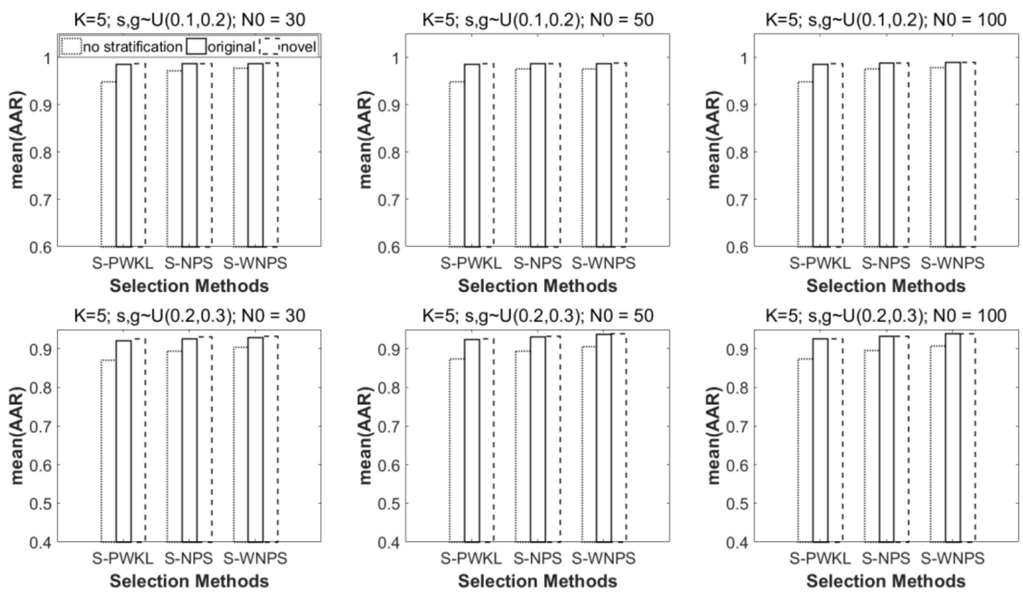

Figure S17. mean of *AAR*s under the DINO model when K = 5 and number of strata = 5

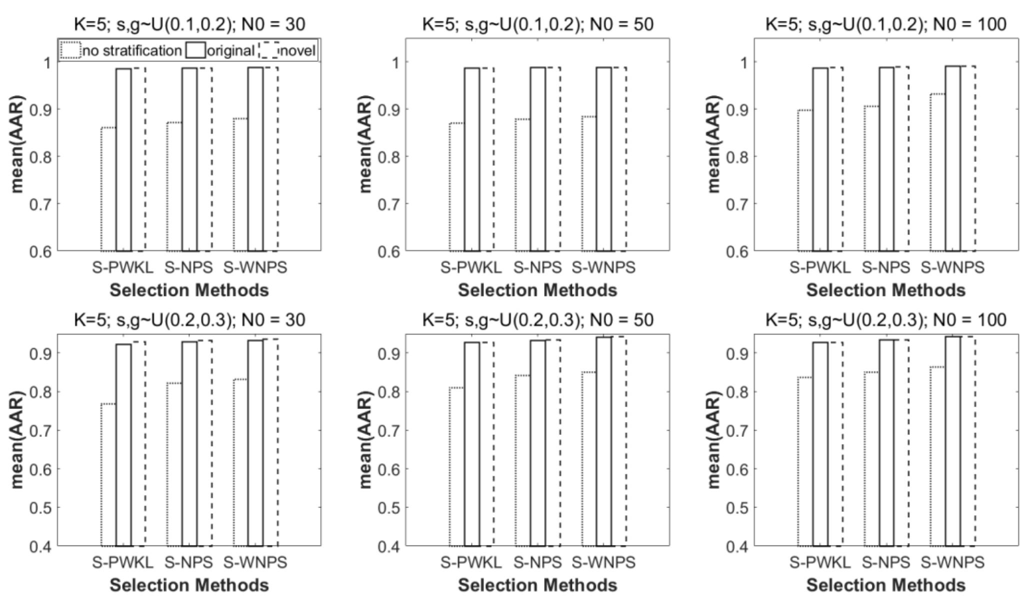

Figure S18. mean of *AAR*s under the RRUM model when $K = 5$ and number of strata = 5