Online Appendix to Accompany Bauer, D.J. and Sterba, S.K. (Submitted). *Comparing Approaches for Fitting Multilevel Models with Ordinal Outcomes.*

Figure and Table Captions

Figure A. Average bias for the threshold estimates, across estimator, number of outcome categories, and distribution shape. *Notes.* The estimator of linear REML (Restricted normal-theory Maximum Likelihood) was used when fitting the linear multilevel model. The estimators of PQL (Penalized Quasi-Likelihood) or logistic ML (with adaptive quadrature) were used when fitting the logistic multilevel model. Points for two-category conditions are not connected to points for 3-7 category conditions because their distribution shapes do not exactly correspond. Results are collapsed over the number of clusters, cluster size, and the magnitude of the random effects.

Figure B. Average bias for the threshold estimates, across logistic estimators, number of outcome categories, and cluster size. *Notes.* Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

Figure C. Bias for the fixed effect estimate of the level 1 predictor X_{ij}, across estimator, number of outcome categories, and distribution shape. *Notes.* The estimator of linear REML (Restricted normal-theory Maximum Likelihood) was used when fitting the linear multilevel model. The estimators of PQL (Penalized Quasi-Likelihood) or logistic ML (with adaptive quadrature) were used when fitting the logistic multilevel model. Results are collapsed over the number of clusters, cluster size, and the magnitude of the random effects.

Figure D. Bias for the fixed effect estimate of the level 2 predictor W_j, across estimator, number of outcome categories, and distribution shape. *Notes.* The estimator of linear REML (Restricted normal-theory Maximum Likelihood) was used when fitting the linear multilevel model. The estimators of PQL (Penalized Quasi-Likelihood) or logistic ML (with adaptive quadrature) were used when fitting the logistic multilevel model. Results are collapsed over the number of clusters, cluster size, and the magnitude of the random effects.

Figure E. Bias for the fixed effect estimate of the cross-level interaction $X_{ij}W_j$, across estimator, number of outcome categories, and distribution shape. *Notes.* The estimator of linear REML (Restricted normal-theory Maximum Likelihood) was used when fitting the linear multilevel model. The estimators of PQL (Penalized Quasi-Likelihood) or logistic ML (with adaptive quadrature) were used when fitting the logistic multilevel model. Results are collapsed over the number of clusters, cluster size, and the magnitude of the random effects.

Figure F. Bias for the fixed effect estimate of the level 1 predictor X_{ij}, across logistic estimators, number of outcome categories, and cluster size. *Notes.* Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

Figure G. Bias for the fixed effect estimate of the level 2 predictor W_j, across logistic estimators, number of outcome categories, and cluster size. *Notes.* Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

Figure H. Bias for the fixed effect estimate of the cross-level interaction $X_{ij}W_j$, across logistic estimators, number of outcome categories, and cluster size. *Notes.* Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

Figure I. Bias for the fixed effect SE of the level 1 predictor X_{ij}, across logistic estimators, number of outcome categories, and cluster size. *Notes.* SE=standard error. Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

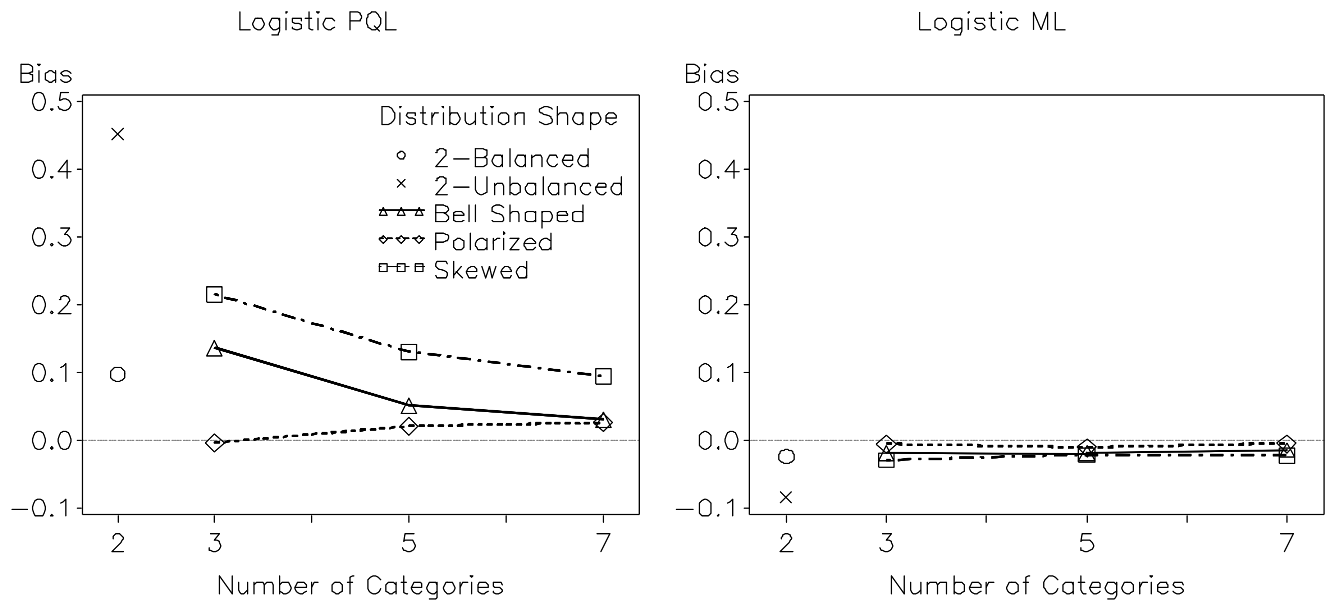
Figure J. Bias for the fixed effect SE of the level 2 predictor W_j, across logistic estimators, number of outcome categories, and cluster size. *Notes.* SE=standard error. Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

Figure K. Bias for the fixed effect SE of the cross-level interaction $X_{ij}W_j$, across logistic estimators, number of outcome categories, and cluster size. *Notes.* SE=standard error. Logistic estimators were either PQL (Penalized Quasi-Likelihood) or logistic ML (Maximum Likelihood) with adaptive quadrature. Results are collapsed over number of clusters and distribution shape and do not include linear multilevel model conditions.

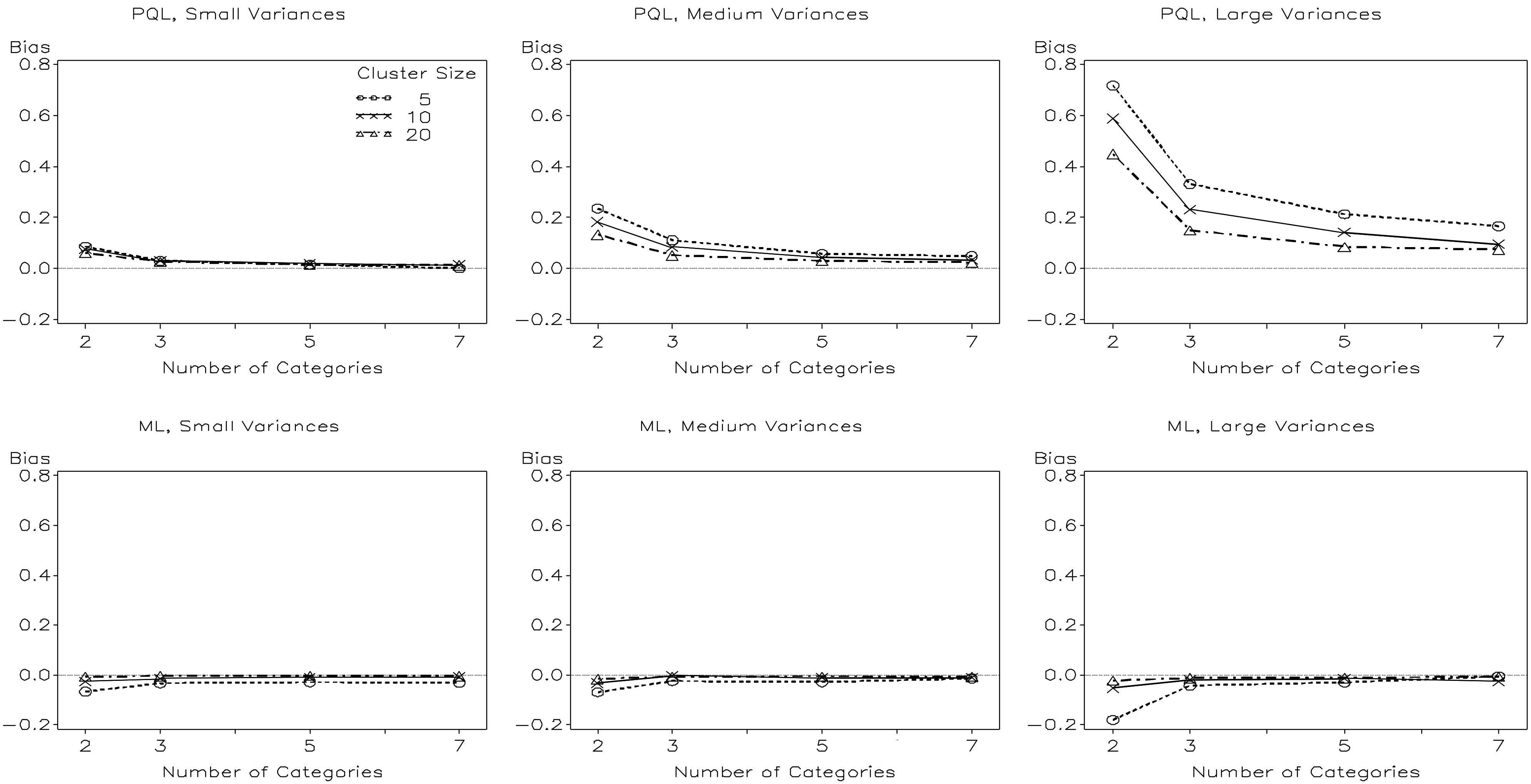
Table A. Coverage for fixed effects (averaged across fixed effects for X_{ij}, W_j, $X_{ij}W_j$) across number of categories, cluster size, variance component sizes, and logistic estimators. *Notes.* PQL=Penalized Quasi-Likelihood and logistic ML=Maximum Likelihood with adaptive quadrature

Table B. Coverage for fixed effects (averaged across fixed effects for X_{ij}, W_j, $X_{ij}W_j$), across number of categories, distribution shape and estimator. *Notes.* PQL=Penalized Quasi-Likelihood and logistic ML=Maximum Likelihood with adaptive quadrature and REML=Restricted normal-theory Maximum Likelihood. 2-bal=binary balanced condition; 2-unbal=binary unbalanced condition.

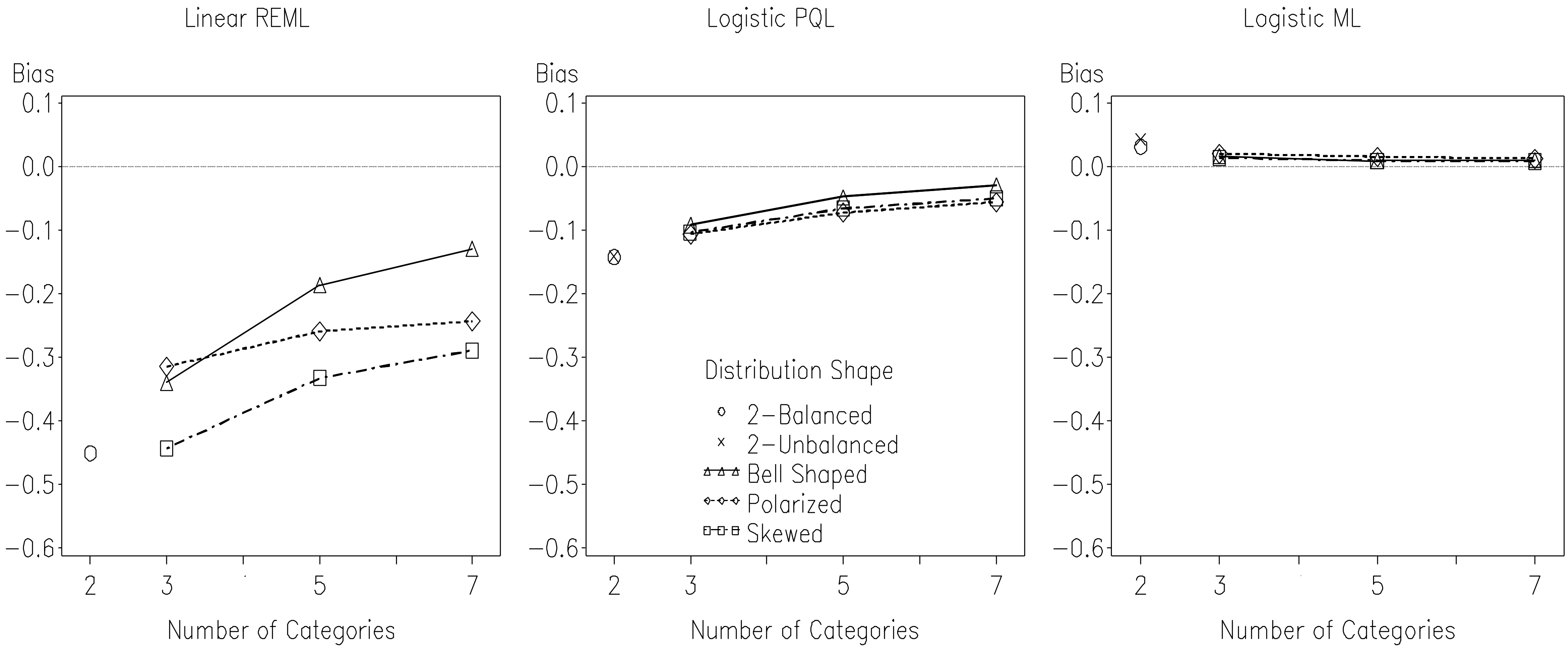
Online Appendix Figure A



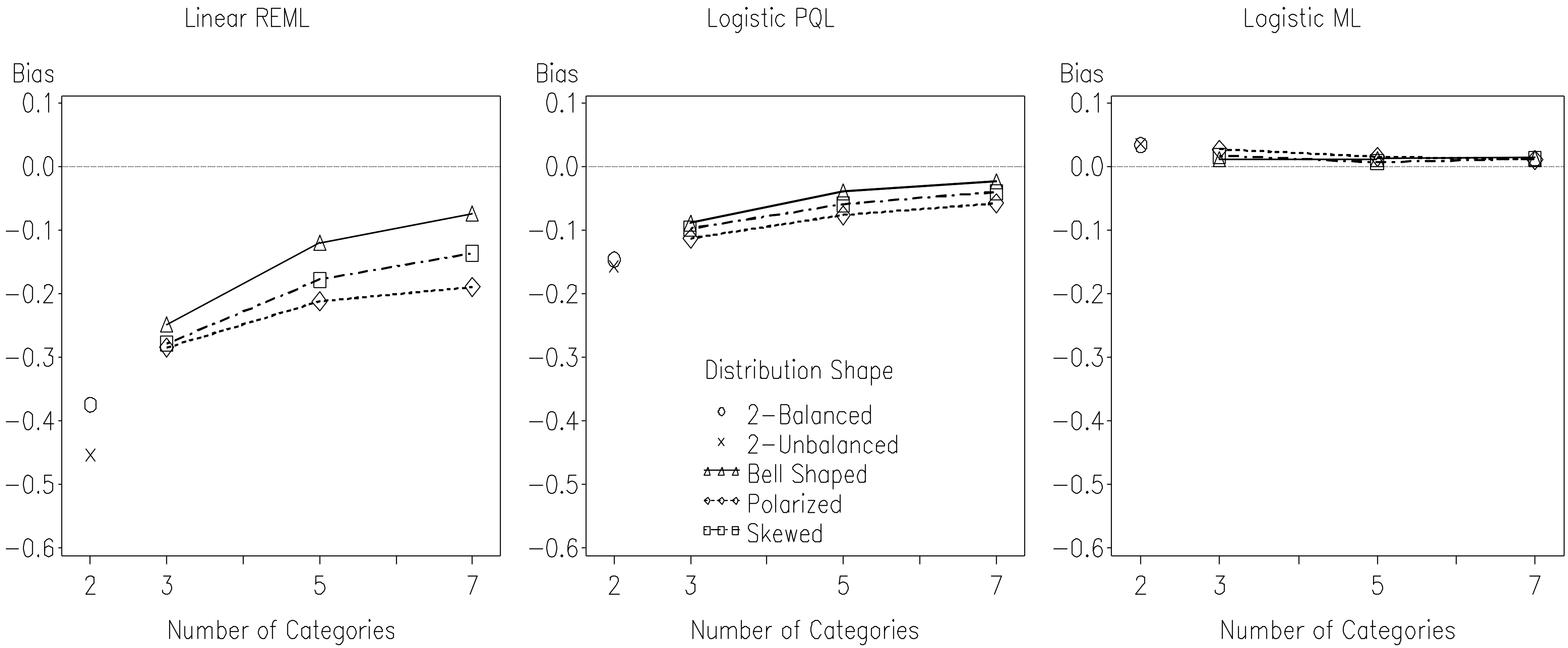
Online Appendix Figure B



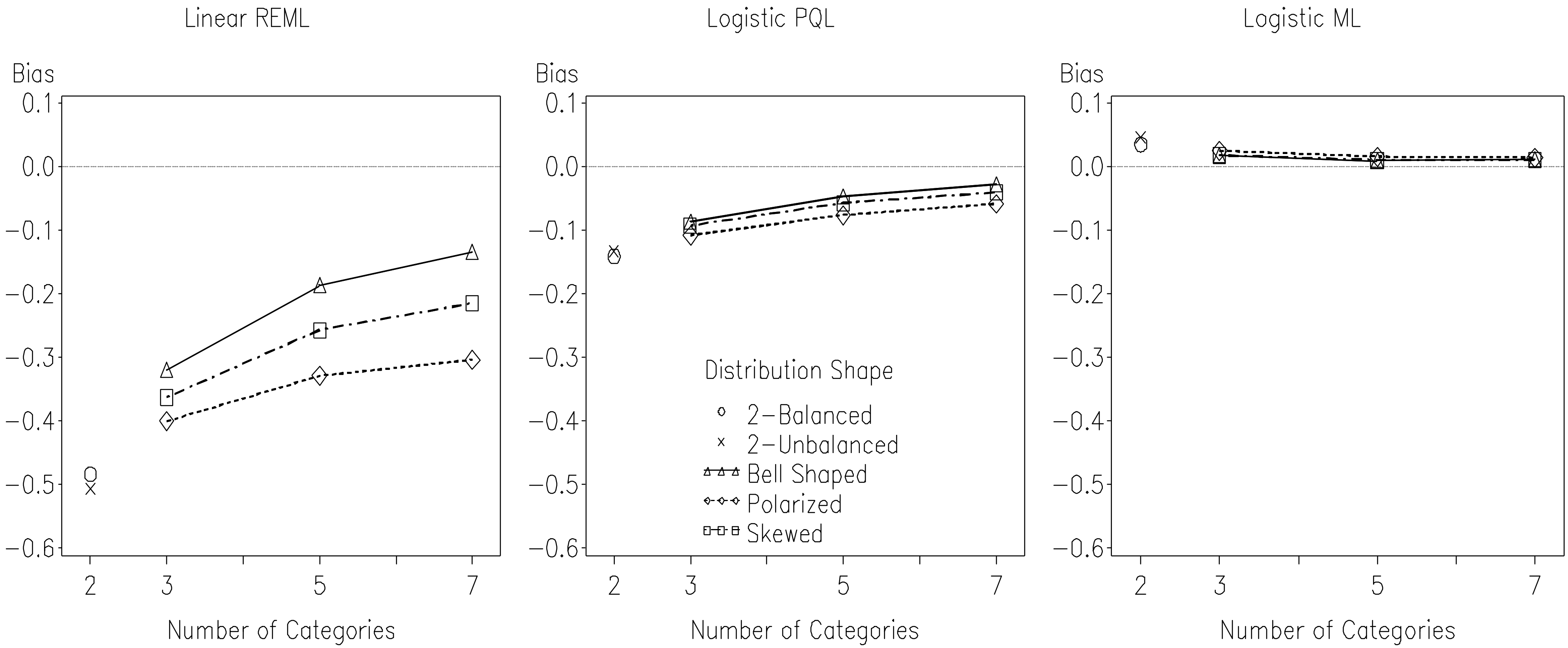
Online Appendix Figure C



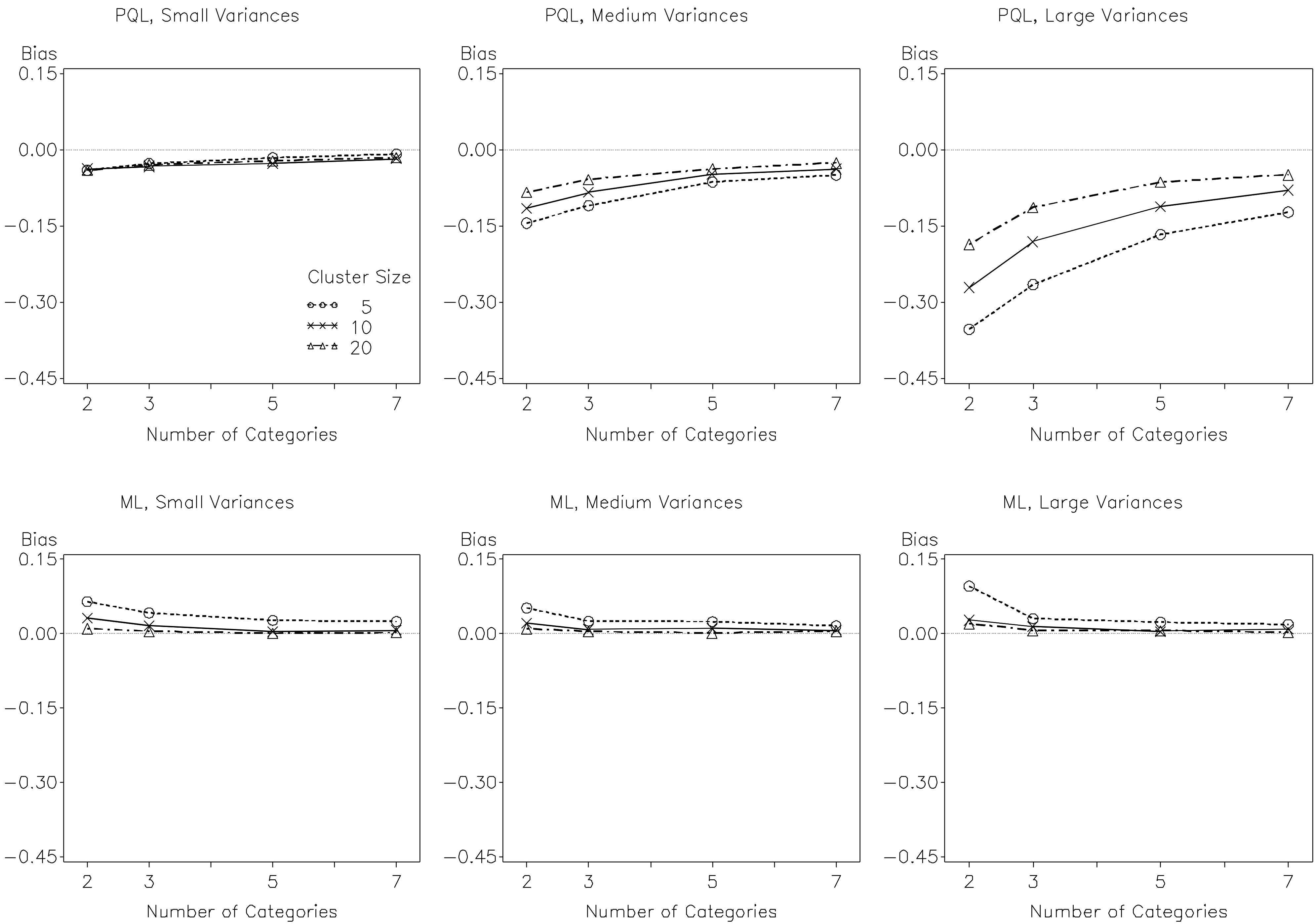
Online Appendix Figure D



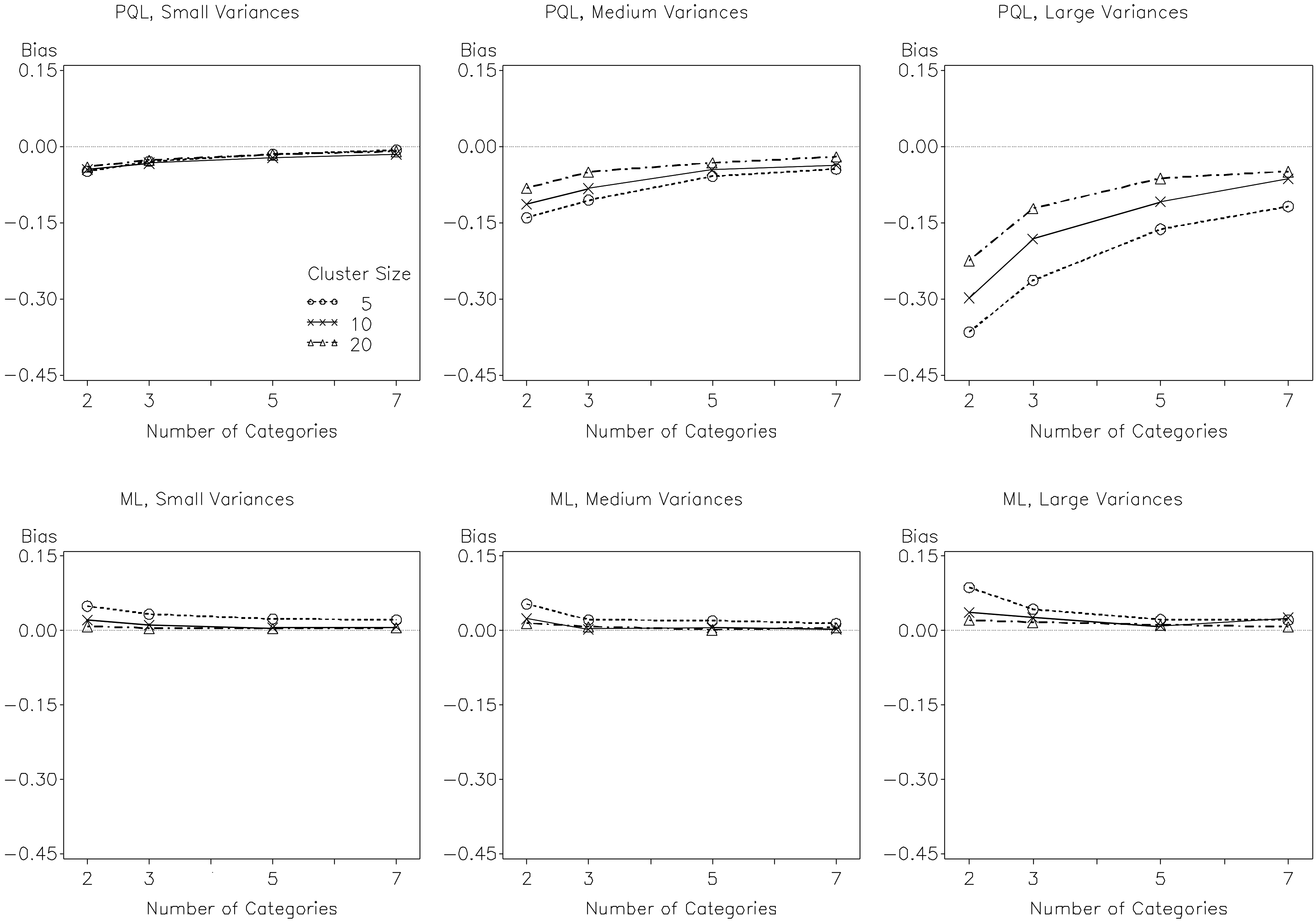
Online Appendix Figure E



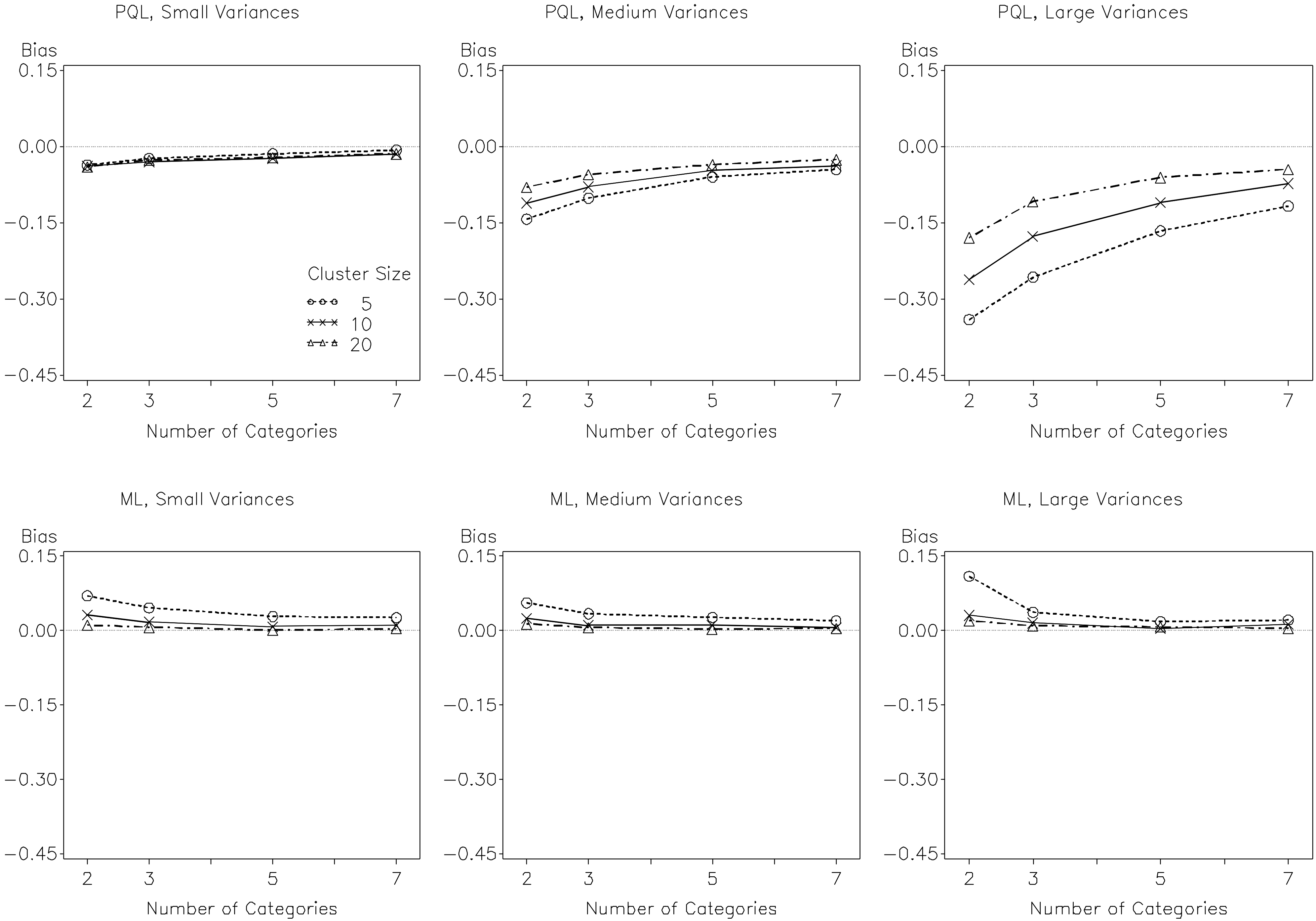
Online Appendix Figure F



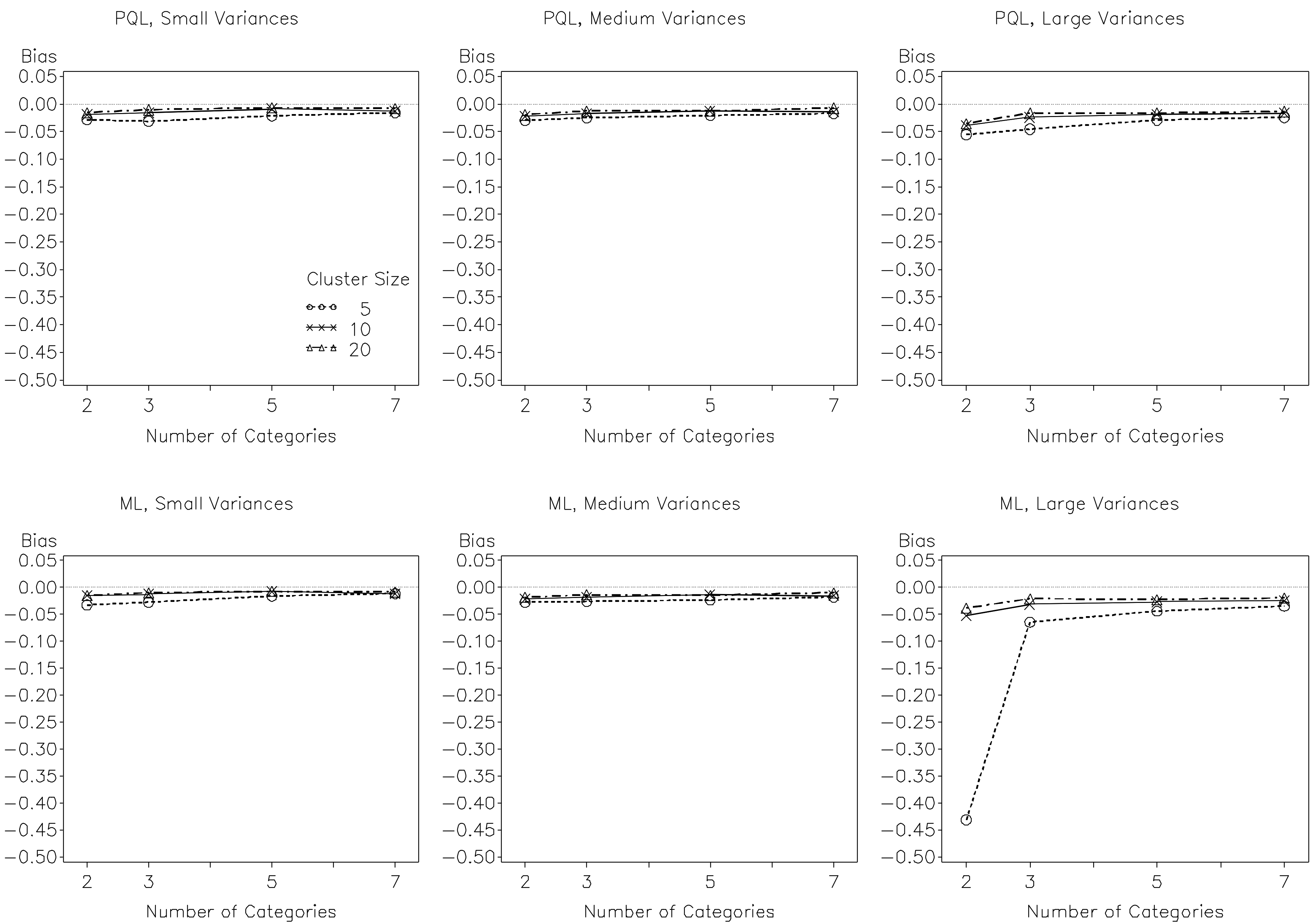
Online Appendix Figure G



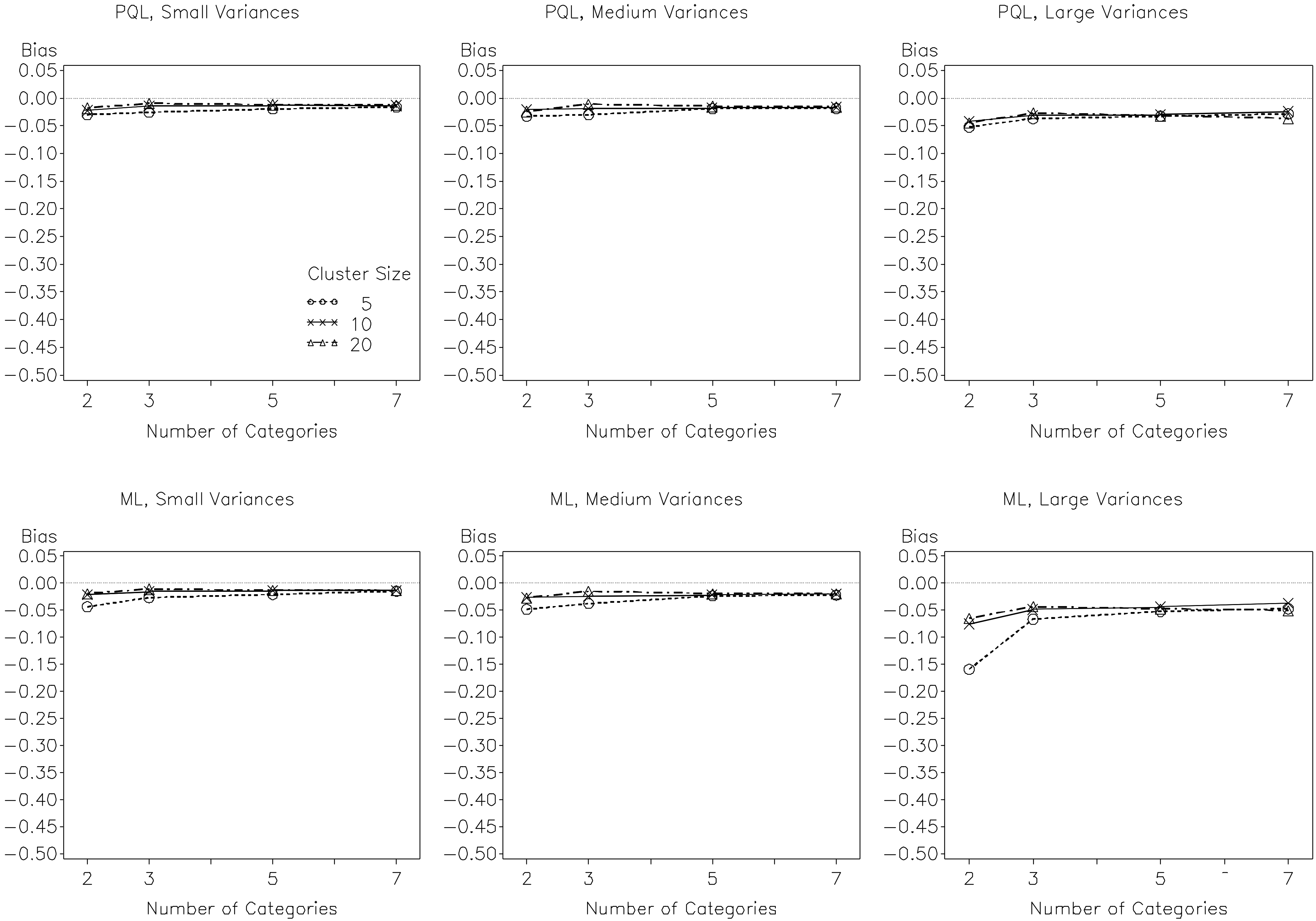
Online Appendix Figure H



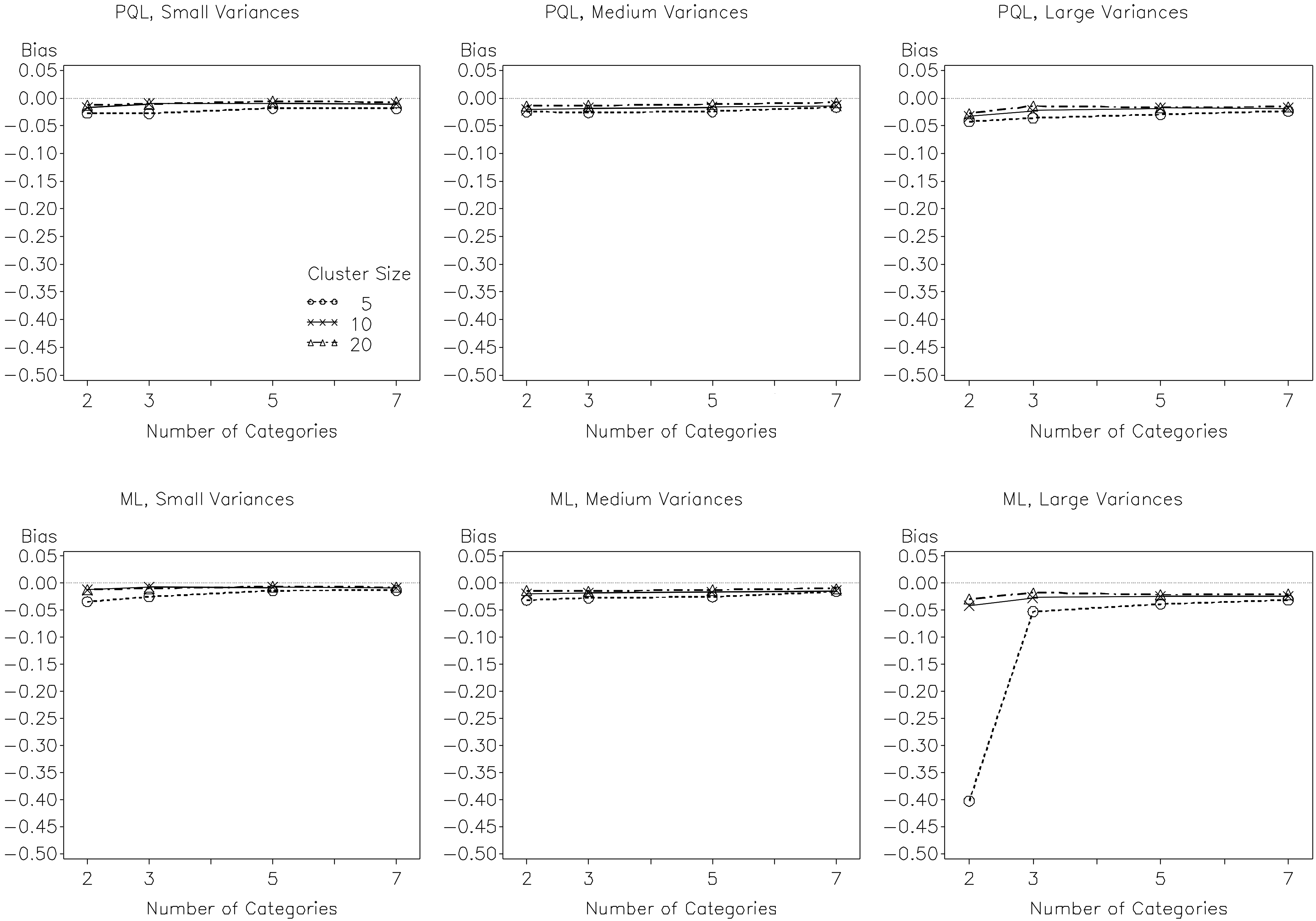
Online Appendix Figure I



Online Appendix Figure J



Online Appendix Figure K



Online Appendix Table A.

	PQL								
	Small Variances			Medium Variances			Large Variances		
	cluster size			cluster size			cluster size		
categories	5	10	20	5	10	20	5	10	20
2	.93	.92	.91	.83	.83	.86	.56	.64	.76
3	.93	.93	.93	.86	.87	.89	.68	.79	.88
5	.94	.93	.94	.90	.91	.92	.83	.89	.92
7	.95	.94	.94	.92	.92	.94	.89	.92	.93

	Logistic ML								
	Small Variances			Medium Variances			Large Variances		
	cluster size			cluster size			cluster size		
categories	5	10	20	5	10	20	5	10	20
2	.97	.95	.95	.96	.95	.95	.96	.95	.94
3	.95	.95	.95	.95	.95	.94	.95	.95	.95
5	.96	.95	.95	.95	.95	.94	.95	.95	.94
7	.96	.95	.94	.95	.95	.95	.95	.95	.94

Online Appendix Table B.

	Linear REML					PQL					Logistic ML				
	distribution shape					distribution shape					distribution shape				
categories	2-bal	2-unbal	bell	polarized	skew	2-bal	2-unbal	bell	polarized	skew	2-bal	2-unbal	bell	polarized	skew
2	.18	.12				.79	.81				.95	.95			
3			.36	.29	.31			.87	.87	.84			.95	.95	.95
5			.65	.47	.44			.93	.91	.89			.95	.95	.95
7			.78	.55	.48			.94	.93	.91			.95	.95	.95

Details on PQL Estimation

Here we provide additional details on the PQL estimator considered in the simulation study. We begin by showing this estimator for a binary outcome then note the generalization for ordinal outcomes.

For a binary outcome *Y*, the multilevel logistic model can be expressed as

$$Y_{ij} = g^{-1}\left(\eta_{ij}\right) + r_{ij} \tag{1}$$

where $g^{-1}\left(\eta_{ij}\right)$ is the inverse of the logistic link function, returning the expected value (predicted probability) for Y_{ij}:

$$g^{-1}\left(\eta_{ij}\right) = \frac{\exp\left(\eta_{ij}\right)}{1+\exp\left(\eta_{ij}\right)} \tag{2}$$

The model is "linearized" using a Taylor series expansion of $g^{-1}\left(\eta_{ij}\right)$, with the linearization improved across successive iterations (see Raudenbush & Bryk, 2001, p. 457-459 and Goldstein, 1995, p. 112-113). Using a first-order expansion about the estimate of η_{ij} obtained in iteration *s*, denoted $\eta_{ij}^{(s)}$, we may write

$$g^{-1}\left(\eta_{ij}\right) \approx g^{-1}\left(\eta_{ij}^{(s)}\right) + g^{-1\prime}\left(\eta_{ij}\right)\left(\eta_{ij} - \eta_{ij}^{(s)}\right) \tag{3}$$

where $g^{-1\prime}\left(\eta_{ij}\right)$ is the first derivative of the inverse link function

$$g^{-1\prime}\left(\eta_{ij}\right) = \frac{\exp\left(\eta_{ij}\right)}{\left[\exp\left(\eta_{ij}\right)+1\right]^{2}} \tag{4}$$

The function $g^{-1\prime}\left(\eta_{ij}\right)$ is evaluated at the current estimates $\eta_{ij}^{(s)}$ to produce the value $w_{ij}^{(s)}$:

$$w_{ij}^{(s)} = \frac{\exp\left(\eta_{ij}^{(s)}\right)}{\left[\exp\left(\eta_{ij}^{(s)}\right)+1\right]^{2}} \tag{5}$$

The linear approximation in Equation (3), using $w_{ij}^{(s)}$ in place of $g^{-1\prime}\left(\eta_{ij}\right)$, is then substituted into Equation (1) to yield

$$Y_{ij} = g^{-1}\left(\eta_{ij}^{(s)}\right) + w_{ij}^{(s)}\left(\eta_{ij} - \eta_{ij}^{(s)}\right) + r_{ij} \tag{6}$$

Now the equation can be rearranged so that the right side is linear:

$$\frac{Y_{ij} - g^{-1}\left(\eta_{ij}^{(s)}\right)}{w_{ij}^{(s)}} + \eta_{ij}^{(s)} = \eta_{ij} + \frac{r_{ij}}{w_{ij}^{(s)}} \tag{7}$$

The quantities Y_{ij}, $g^{-1}\left(\eta_{ij}^{(s)}\right)$, $w_{ij}^{(s)}$, and $\eta_{ij}^{(s)}$ are all "known" in the sense that they are either observed or computable at each iteration *s*; hence, we can rewrite Equation (7) as

$$Z_{ij} = \eta_{ij} + e_{ij} \tag{8}$$

where

$$Z_{ij} = \frac{Y_{ij} - g^{-1}\left(\eta_{ij}^{(s)}\right)}{w_{ij}^{(s)}} + \eta_{ij}^{(s)} \tag{9}$$

is a *working variate* or *pseudo-data*, updated from one iteration to the next and $e_{ij} = r_{ij} / w_{ij}^{(s)}$ is the Level-1 residual term scaled by $w_{ij}^{(s)}$. Note that the model for the working variate is linear with heteroscedastic residuals. We assume that a normal distribution will hold, at least approximately, for the rescaled residuals, or $e_{ij} \sim N\left(0, 1/w_{ij}^{(s)}\right)$. Note that $w_{ij}^{(s)}$ effectively serves as a weight for the observation (similar to weighted least squares), allowing for heteroscedasticity of the response distribution. With this approximation in place the model for the working variate is of the same form as the linear multilevel model. That is, the working variate is a linear combination of normally distributed random effects and normally distributed residuals. The usual normal-theory estimators used for linear multilevel models (e.g., Restricted Maximum Likelihood, REML) can therefore be applied to fit the model in Equation (8).

In practice the algorithm is started by choosing initial values for $\eta_{ij}^{(0)}$ and $Z_{ij}^{(0)}$. For instance, initial values might be set to $\eta_{ij}^{(0)} = \text{logit}(.25)$ if $Y_{ij} = 0$ and $\eta_{ij}^{(0)} = \text{logit}(.75)$ if $Y_{ij} = 1$. The model in Equation (8) is then fit to $Z_{ij}^{(0)}$ (e.g., by REML), using $w_{ij}^{(0)}$ as an observation weight. The results are used to generate $\eta_{ij}^{(1)}$ and $Z_{ij}^{(1)}$, and the process is repeated until the parameter estimates no longer change.

For ordinal data, several additional considerations apply. First, the model is not fit directly to the ordinal variable, but rather to a vector of cumulative binary coding variables representing the ordinal scores. That is, for an ordinal variable *Y* coded with categories $c = 1, 2, \ldots, C$, $C - 1$ binary variables are created such that $Y_{ij}^{(c)} = 1$ if $Y_{ij} \le c$, with the last category, *C*, omitted. The equations provided above still apply, but with $Y_{ij}^{(c)}$ as the referent dependent variable. The inverse link in Equation (2) then returns the expected value of $Y_{ij}^{(c)}$, which is the cumulative probability $P\left(Y_{ij} \le c\right) = \varphi_{ij}^{(c)}$. Second, the linear predictor within the inverse link function must be augmented to include the threshold parameters. That is, in the above equations, η_{ij} is replaced by $\nu^{(c)} - \eta_{ij}$ where $\nu^{(c)}$ is a threshold parameter. Third, the $C - 1$ binary values constructed to represent each ordinal score Y_{ij} are not independent. Specifically, $COV\left(Y_{ij}^{(c)}, Y_{ij}^{(c')}\right) = \varphi_{ij}^{(c)}\left(1 - \varphi_{ij}^{(c')}\right)$ for $c \le c'$ (McCullagh & Nelder, 1989, p. 167). The working variate scores must therefore be treated as correlated when fitting the linearized model.