Bootstrapping data arrays of arbitrary order

Art B. Owen
Stanford University

Dean Eckles
Stanford University and Facebook Inc.

June 2011

Abstract

In this paper we study a bootstrap strategy for estimating the variance of a mean taken over large multifactor crossed random effects data sets. We apply bootstrap reweighting independently to the levels of each factor, giving each observation the product of its factor weights. No exact bootstrap exists for this problem (McCullagh, 2000). We show that the proposed bootstrap is mildly conservative, under sufficient conditions that allow very unbalanced and heteroscedastic inputs. Earlier results for a resampling bootstrap only apply to two factors and are not suitable to online computation. The proposed reweighting approach can be implemented in parallel and online settings. The results for this method apply to any number of factors. The method is illustrated using a 3 factor data set of comment lengths from Facebook.

Keywords: Bayesian pigeonhole bootstrap, online bagging, online bootstrap, relational data, tensor data, unbalanced random effects

1 Introduction

In IID sampling, the bootstrap provides reliable variance estimates and confidence intervals under very weak assumptions on the mechanism generating our data. But many data sets have no IID structure for us to draw on. For example, with the famous Netflix data (Bennett and Lanning, 2007) multiple ratings from the same viewer should be modeled as dependent. Similarly multiple ratings on the same movie are dependent. Neither rows nor columns are IID, and a crossed random effects model with interactions is a more reasonable structure. Large crossed data sets of two or even more factors are commonplace in electronic commerce and we might expect them to appear in other settings where data collection is automated.

The crossed random effects setting is challenging for inference. Methods in Searle et al. (1992) apply for Gaussian random effects models, but for large unbalanced cases, the necessary linear algebra becomes prohibitively expensive. We might therefore turn to resampling. But McCullagh (2000) proved that there does not exist a bootstrap algorithm which will correctly estimate the

variance of the overall average of a crossed random effects model, even in the case of perfectly balanced data (no missing values).

Large sparse data sets with crossed random effects commonly arise. There can easily be more than two factors. For example, at an online bookstore, customer C might arrive from IP address I, enter query Q, read review R about book B, buy that book with credit card K at time T, and have it shipped to address A. The data logs then have a sparse table of observed (C, I, Q, R, B, K, T, A) values. Crossed random effects with various interactions are a natural model for such data sets.

These data sets may be extremely sparse and unbalanced. The Netflix data have about 1% of observations present and the number of ratings per movie has an enormous range as does the number of ratings per customer. When there are more than 2 factors, the data may be much sparser still.

There does exist an approximate bootstrap method which is easy to apply in the two factor case. In it, one takes bootstrap samples of the row entities and, independently, of the column entities. If, for example, row i is sampled A times and column j is sampled B times then the ij observation will appear AB times in the bootstrap sample. This bootstrap has been used in item-response theory by Brennan et al. (1987) and Wiley (2001) to study variance components in educational test data. They noticed that bootstrap methods were biased.

After proving the non-existence of an unbiased bootstrap, McCullagh (2000) shows that bootstrapping rows and columns independently, gives a mildly conservative variance estimate for balanced crossed random effects data with homoscedastic errors. Owen (2007) shows that this bootstrap remains conservative (and usually mildly so) for sparse and unbalanced crossed random effects. That framework allows every row and column (e.g. customer and movie) and even every interaction to have its own variance. Resampling is then reliable and it spares the user from having to estimate all of those variances.

Methods that reweight data via IID random weights (Rubin, 1981; Newton and Raftery, 1994) are an appealing alternative to resampling, which amounts to using multinomial weights. First, it is simpler to apply reweighting to large scale parallelized computations, as researchers in online bootstrapping (Oza, 2001; Lee and Clyde, 2004) do. The reason is that large data sets are stored in a distributed fashion and then multinomial sampling brings substantial communication and synchronization costs. Second, resampling simplifies variance expressions by avoiding the negative dependence from the multinomial distribution. This makes it easier to develop expressions for problems with more than two factors. Many statistics of interest can be extended to weighted samples. For those that cannot, weights that take the form of nonnegative integers can be implemented by counting multiple copies of the corresponding observations.

Using notation and approximations defined below, the main facts are as follows. For $r = 2$ factors, the random effects variance of a sample mean takes the form $(\nu_{\{1\}}\sigma^2_{\{1\}} + \nu_{\{2\}}\sigma^2_{\{2\}} + \sigma^2_{\{1,2\}})/N$ for N observations. The subscripted ν quantities are easily computable from the data while the subscripted σ^2's are variance components. Naive bootstrapping produces an estimate close to

$(\sigma^2_{\{1\}} + \sigma^2_{\{2\}} + \sigma^2_{\{1,2\}})/N$ which is grossly inadequate because it turns out that often $1 \ll \nu_{\{j\}} \ll N$. Resampling both rows and columns leads to a variance estimate close to $((\nu_{\{1\}}+2)\sigma^2_{\{1\}} + (\nu_{\{2\}}+2)\sigma^2_{\{2\}} + 3\sigma^2_{\{1,2\}})/N$, which is mildly conservative when $\nu_{\{j\}} \gg 1$ and the σ's are of comparable magnitude. It is up to three times as large as it should be in the event that $\sigma^2_{\{j\}} \ll \sigma^2_{\{1,2\}}$. For the Netflix data, a naive bootstrap variance can be too small by as much as roughly 56,200, a far more serious error than overestimating by a factor of 3.

Our main contributions are:

1) showing that the same stylized facts hold for reweighting $r = 2$ factors,

2) generalizing the reweighting results to $r \geq 2$ factors, and

3) analyzing the heteroscedastic random effects case.

As an example, the $3\sigma^2_{\{1,2\}}$ from $r = 2$ becomes $(2^r - 1)\sigma^2_{\{1,2,\ldots,r\}}$ and we find expressions for all $2^r - 1$ coefficients. Under reasonable conditions, for which we note exceptions, this bootstrap magnifies a k-factor variance component by roughly $2^k - 1$. Under simply described conditions, the $k = 1$ terms dominate the variance and then the variance magnification becomes negligible.

An outline of the paper is as follows. Section 2 introduces our notation for the random effects model and some observation counts and then defines the random effects variance that we seek to estimate. Section 3 considers naive bootstrap methods that simply resample or reweight the observations as if they were IID. They seriously underestimate the true variance unless the only nonzero variance component is that of the highest order interaction. Reweighting has a slight advantage because it allows one to step up the sampling variance to compensate for cases where the naive bootstrap variance is only a modest underestimate. Section 4 introduces a factorial reweighting bootstrap strategy. For data with $r = 2$ factors the reweighting results closely match the resampling results from Owen (2007). This section includes an interpretable approximation to the exact bootstrap variance. Section 6 considers the heteroscedastic case, where every variance component at every combination of its factors has its own variance parameter. When the main effects are dominant, then the proposed bootstrap closely matches the desired variance even in the heteroscedastic setting. Section 7 describes repeated observations and factors nested inside the ones being reweighted. Section 9 has a numerical example from Facebook. In that dataset, UK-based users make longer comments than do US-based users, when posting from mobile devices. The reverse holds for comments made at Facebook's standard interface to the web. The differences are small, but statistically significant, even after taking account of a three factor structure (commenter, sharer, and URL). Section 8 briefly sketches how reliable variance estimation for a mean extends to other problems. The proofs appear in an Appendix.

2 Notation and random effects model

The random variables of interest take the form $X_{i_1,i_2,\cdots,i_r} \in \mathbb{R}^d$ for integers $i_j \geq 1$ and $j = 1, \ldots, r$. To simplify notation, we write $X_{\boldsymbol{i}}$ for $\boldsymbol{i} = (i_1, \ldots, i_r)$. We work with dimension $d = 1$. The generalization to $d \geq 1$ is straightforward.

We have in mind applications where each value of i_j corresponds to one level of a categorical variable with many potential values. In internet applications, index values i_j might represent users, URLs, IP addresses, ads, query strings and so on. There may be no a priori upper bound on the number of distinct levels for i_j.

The data are composed of N of these random variables, where $1 \le N < \infty$. The binary variable $Z_{\boldsymbol{i}}$ takes the value 1 when observation $X_{\boldsymbol{i}}$ is present and $Z_{\boldsymbol{i}} = 0$ when $X_{\boldsymbol{i}}$ is absent. We work conditionally on $Z_{\boldsymbol{i}}$ so that they are nonrandom. In practice the pattern of missingness in $Z_{\boldsymbol{i}}$ may be important. We avoid modeling $Z_{\boldsymbol{i}}$ in order to focus on estimating variance, apart from some brief remarks in Section 8.

The letters u and v denote subsets of $[r] \equiv \{1, \dots, r\}$ throughout. The summation $\sum_u$ is taken over all 2^r subsets of $[r]$, and other summations, such as $\sum_{v \supseteq u}$ denote sums over the first named set (here v) subject to the indicated condition with the other set(s) (here u) held fixed. The index i_u extracts the components i_j for $j \in u$. Then $i_u = i'_u$ means that $i_j = i'_j$ for all $j \in u$.

Our ***r*-fold crossed random effects model** is

$$X_{\boldsymbol{i}} = \mu + \sum_{u \neq \varnothing} \varepsilon_{\boldsymbol{i},u} \tag{1}$$

where $\mu \in \mathbb{R}$ and $\varepsilon_{\boldsymbol{i},u}$ are mean 0 random variables that depend on $\boldsymbol{i}$ only through i_u. We have $\varepsilon_{\boldsymbol{i},u} = \varepsilon_{\boldsymbol{i}',u}$ if $i_u = i'_u$ and $\varepsilon_{\boldsymbol{i},u}$ independent of $\varepsilon_{\boldsymbol{i}',u}$ otherwise. The covariance of $\varepsilon_{\boldsymbol{i},u}$ and $\varepsilon_{\boldsymbol{i}',u'}$ is

$$\mathrm{Cov}(\varepsilon_{\boldsymbol{i},u}, \varepsilon_{\boldsymbol{i}',u'}) = \mathbb{E}(\varepsilon_{\boldsymbol{i},u}\varepsilon_{\boldsymbol{i}',u'}) = \sigma_u^2 \mathbf{1}_{u=u'} \mathbf{1}_{i_u = i'_u} \tag{2}$$

for $\sigma_u^2 < \infty$.

The sample mean of X is the ratio

$$\bar{X} = \sum_{\boldsymbol{i}} X_{\boldsymbol{i}} Z_{\boldsymbol{i}} \Big/ \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}, \tag{3}$$

where the sums are over all index values $\boldsymbol{i}$. The denominator in (3) is the total number N of observations. Our goal is to estimate the variance of $\bar{X}$ by resampling methods.

2.1 Partial duplicate observations

We will need to keep track of the extent to which different observations have the same index values, in order to properly reflect correlations among the $X_{\boldsymbol{i}}$.

For each $\boldsymbol{i}$ and $u \subseteq [r]$, the number

$$N_{\boldsymbol{i},u} = \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'} \mathbf{1}_{i_u = i'_u}$$

counts how many observations match $X_{\boldsymbol{i}}$ for all indices $j \in u$. If $Z_{\boldsymbol{i}} = 1$ then $N_{\boldsymbol{i},u} \ge 1$ because $X_{\boldsymbol{i}}$ matches itself. By convention $N_{\boldsymbol{i},\varnothing} = N$ and $N_{\boldsymbol{i},[r]} = 1$.

The quantity

$$\nu_u = \frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} N_{\boldsymbol{i},u} \geq 1$$

is the average number of matches in the subset u for observations in the data set, and $\nu_{[r]} = 1$.

The most important of the ν_u are for singletons $u = \{j\}$. The value $\nu_{\{j\}}$ has a quadratic dependence on the pattern of duplication in the data. To see this, write $n_{\ell j} = \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} 1_{i_j=\ell}$ for the number of times that variable j is equal to ℓ in the data. Then $\nu_{\{j\}} = N^{-1}\sum_{\ell=1}^{\infty} n_{\ell,j}^2$ because each $N_{\boldsymbol{i},\{j\}} = n_{i_j,j}$ appears $n_{i_j,j}$ times in the summation defining $\nu_{\{j\}}$.

If $u \subseteq v$ then $\nu_u \geq \nu_v$. In some applications $\nu_u \gg \nu_v$ for proper subsets $u \subsetneq v$. For those applications, multiple matches are very unusual. In other settings two factors, say i_1 and i_2 might be highly though not perfectly dependent (e.g., customer ID and phone number) and then $\nu_{\{1,2\}}$ might be only slightly smaller than $\nu_{\{1\}}$ or $\nu_{\{2\}}$. We return to this issue in Section 5.

The specific pair of data values $\boldsymbol{i}$ and $\boldsymbol{i}'$ match in components

$$M_{\boldsymbol{i}\boldsymbol{i}'} = \{j \in [r] \mid i_j = i'_j\}.$$

For the motivating data, most of the $M_{\boldsymbol{i}\boldsymbol{i}'}$ are empty and most of the rest have cardinality $|M_{\boldsymbol{i}\boldsymbol{i}'}| = 1$. We have $|M_{\boldsymbol{i}\boldsymbol{i}'}| = r$ if and only if $\boldsymbol{i} = \boldsymbol{i}'$. Although $M_{\boldsymbol{i}\boldsymbol{i}'}$ is defined for all pairs $\boldsymbol{i}$ and $\boldsymbol{i}'$ we only use it when $Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'} = 1$, that is when both $X_{\boldsymbol{i}}$ and $X_{\boldsymbol{i}'}$ have been observed, and the term 'most' above refers to these pairs.

For each $\boldsymbol{i}$ and $k = 0, 1, \ldots, r$, the number

$$N_{\boldsymbol{i},k} = \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'} \mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}$$

counts how many observations match $X_{\boldsymbol{i}}$ in exactly k places.

2.2 Random effects variance of $\bar{X}$

Here we record the true variance of $\bar{X}$, using the random effects model. This is the quantity we hope to estimate by bootstrapping.

Theorem 1. *In the random effects model* (1)

$$\mathrm{Var}(\bar{X}) = \frac{1}{N}\sum_{u\neq\varnothing} \nu_u \sigma_u^2. \tag{4}$$

The contributions of the variance components σ_u^2 are proportional to the duplication indices ν_u. For large sparse data sets we often find that $1 \ll \nu_u \ll N$ when $0 < |u| < r$.

Our bootstrap approximations to this variance are centered around a quantity $(1/N)\sum_{u\neq\varnothing}\gamma_u\sigma_u^2$ for gain coefficients γ_u that depend on the data configuration and the particular bootstrap method. Ideally we want $\gamma_u = \nu_u$. More realistically, some bootstrap methods are able to get $\gamma_u \geq \nu_u$ with γ_u just barely larger than ν_u for the singletons $u = \{j\}$ which we expect to dominate $\mathrm{Var}(\bar{X})$.

3 Naive bootstrap methods

There are two main ways to bootstrap: resampling (Efron, 1979) and reweighting (Rubin, 1981), with the distinction being that the former uses a multinomial distribution on the data while the latter applies independent random weights to the observations.

Naive bootstrap methods simply resample or reweight the N observations without regard to their factorial structure. That is they use the same bootstrap one might use for IID samples. Here we show that naive bootstrap resampling and reweighting have very similar and very unsatisfactory performance.

3.1 Naive resampling

In the naive bootstrap, all N observations are resampled without replacement. The naive bootstrap variance of $\bar{X}$ converges to $1/N$ times the plugin variance of the N observed $X_{\boldsymbol{i}}$ as the number of resampled data sets tends to infinity. We write

$$\mathrm{Var}_{\mathrm{NB}}(\bar{X}) = \frac{1}{N^2}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}(X_{\boldsymbol{i}} - \bar{X})^2 \tag{5}$$

for this limiting value.

Theorem 2. *Under the random effects model* (1)*, the expected value of the naive bootstrap variance of* $\bar{X}$ *is*

$$\mathbb{E}_{\mathrm{RE}}(\mathrm{Var}_{\mathrm{NB}}(\bar{X})) = \frac{1}{N}\sum_{u\neq\varnothing}\sigma_u^2\Big(1-\frac{\nu_u}{N}\Big). \tag{6}$$

When $r > 1$, the naive bootstrap severely underestimates the coefficients of $\sigma^2_{\{j\}}$. For the Netflix data,

$$\begin{aligned}\mathrm{Var}(\bar{X}) &\doteq \frac{1}{N}(56{,}200\,\sigma^2_{\mathrm{movies}} + 646\,\sigma^2_{\mathrm{raters}} + \sigma^2_{\mathrm{interaction}}), \quad \text{while}\\ \mathrm{Var}_{\mathrm{NB}}(\bar{X}) &\le \frac{1}{N}(\sigma^2_{\mathrm{movies}} + \sigma^2_{\mathrm{raters}} + \sigma^2_{\mathrm{interaction}}),\end{aligned}$$

where $N \doteq 100{,}000{,}000$.

Theorem 2 generalizes Lemma 2 of Owen (2007) which treats naive bootstrap sampling for $r = 2$. We note that Owen (2007, p. 391) has an error: it gives the coefficient of $\sigma^2_{\{1,2\}}$ as $1/N$ where it should be $1/(N-1)$.

3.2 Naive reweighting

In the naive Bayesian bootstrap, all N observations are given random weights which are then normalized. Observation $\boldsymbol{i}$ gets weight $W_{\boldsymbol{i}} \sim G$ independently sampled. We assume that G has mean 1 and variance $\tau^2 < \infty$. Typically $\tau^2 = 1$.

The original Bayesian bootstrap (Rubin, 1981) had $W_{\boldsymbol{i}} \sim \mathrm{Exp}(1)$ but other distributions are useful too. Taking $W_{\boldsymbol{i}} \sim \mathrm{Poi}(1)$ gives a result very similar to the usual bootstrap, and it has integer weights. Independent $\mathrm{Bin}(N, 1/N)$ weights would provide a more exact match, but for large N there is no practical difference between $\mathrm{Bin}(N, 1/N)$ and $\mathrm{Poi}(1)$. See Oza (2001) and Lee and Clyde (2004) for uses of Poisson weights in data mining.

Taking $W_{\boldsymbol{i}} \sim \mathbf{U}\{0, 2\} = (\delta_0 + \delta_2)/2$ also has integer values. The algorithm goes 'double or nothing' independently on all N observations. The nonzero integer values are all equal, so these weights correspond to using a random unweighted subset of the data. Double-or-nothing weighting is then a version of half-sampling methods (McCarthy, 1969) without the constraint on the sum of weights, just as Poisson weighting removes a sum constraint from the original bootstrap.

The choice of weights makes a small difference to the bootstrap performance. See Section 3.3.

Each bootstrap resampled mean takes the form

$$\bar{X}^* = T^*/N^*$$

where $T^* = \sum_{\boldsymbol{i}} W_{\boldsymbol{i}} Z_{\boldsymbol{i}} X_{\boldsymbol{i}}$ and $N^* = \sum_{\boldsymbol{i}} W_{\boldsymbol{i}} Z_{\boldsymbol{i}}$. The bootstrap mean T^*/N^* is a ratio estimator of $\bar{X}$. The asymptotic formula for the variance is

$$\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*) = \frac{1}{N^2}\mathbb{E}_{\mathrm{NBB}}\big(\big(T^* - \bar{X}N^*\big)^2\big).$$

The tilde on $\mathrm{Var}_{\mathrm{NBB}}$ is a reminder that this formula is a delta method approximation: it is the variance of a Taylor approximation to $\bar{X}^*$. Because N is usually very large in the target applications, we consider $\widetilde{\mathrm{Var}}_{\mathrm{NBB}}$ to be a reliable proxy for $\mathrm{Var}_{\mathrm{NBB}}$.

Theorem 3. *In the random effects model* (1)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*)\big) = \frac{\tau^2}{N}\sum_{u\neq\varnothing}\sigma_u^2\Big(1 - \frac{\nu_u}{N}\Big). \tag{7}$$

The naive Bayesian bootstrap using $\tau^2 = 1$ has the same average variance as the naive bootstrap. In large data sets we may find that $\nu_u \gg \tau^2(1-\nu_u/N)$ and then the Bayesian bootstrap greatly underestimates the true variance. When $\max_{u\neq\varnothing}\nu_u$ is not too large, then Theorem 3 offers a way to counter this problem. We can take $\tau^2 = \max_{u\neq\varnothing}\nu_u$ and get conservative variance estimates from the naive Bayesian bootstrap. The largest ν_u comes from $u = \{j\}$ for some $j \in [r]$ and it is an easy quantity to compute.

3.3 Bootstrap stability

Any distribution on weights with $\mathbb{E}(W) = 1$ and $\mathrm{Var}(W) = \tau^2$ will have the same value for $\mathbb{E}_{\mathrm{RE}}(\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*))$. Other things being equal, we prefer a more

stable bootstrap. By stability we mean efficiency with respect to bootstrap sampling, as distinct from statistical efficiency with respect to sampling of the original data $X_{\boldsymbol{i}}$. We find that the stability of reweighting depends on the kurtosis $\kappa = \mathbb{E}((W-1)^4)/\tau^4 - 3$ of the weights. The dependence is quite weak unless the $X_{\boldsymbol{i}}$ have a high sample kurtosis $\kappa_x = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}(X_{\boldsymbol{i}} - \bar{X})^4/\sigma^4 - 3$ where $\sigma^2 = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}(X_{\boldsymbol{i}} - \bar{X})^2$.

If we hold the observations $X_{\boldsymbol{i}}$ fixed and implement the bootstrap, doing some number B of replicates, we will estimate the quantity

$$\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*) = \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \mathbb{E}_{\mathrm{NBB}}(W_{\boldsymbol{i}} W_{\boldsymbol{i}'}) Y_{\boldsymbol{i}} Y_{\boldsymbol{i}'}$$

where $Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} - \bar{X}$.

To estimate this variance we may use

$$\begin{aligned}\widehat{\widetilde{\mathrm{Var}}}_{\mathrm{NBB}}(\bar{X}^*) &= \frac{1}{BN^2}\sum_{b=1}^{B}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} W_{\boldsymbol{i},b} W_{\boldsymbol{i}',b} Y_{\boldsymbol{i}} Y_{\boldsymbol{i}'} \\ &= \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i},b}(X_{\boldsymbol{i}} - \bar{X})\right)^2\end{aligned} \tag{8}$$

where $W_{\boldsymbol{i},b}$ are independent identically distributed random weights and $b = 1, \ldots, B$ indexes the bootstrap replications. The hat in (8) represents estimation from B bootstrap samples. It is possible to use (8) with $B = 1$. That such a "unistrap" is possible, reflects the use of a delta method approximation.

Equation (8) is not the usual estimator. The more usual variance estimate is

$$s^2_{\mathrm{NBB}}(\bar{X}^*) = \frac{1}{B-1}\sum_{b=1}^{B}(\bar{X}^*_b - \bar{X}^*_\bullet)^2 \tag{9}$$

where

$$\bar{X}^*_b = \frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i},b} X_{\boldsymbol{i}}, \quad \text{and} \quad \bar{X}^*_\bullet = \frac{1}{B}\sum_{b=1}^{B}\bar{X}^*_b. \tag{10}$$

Theorem 4. *Let W and $W_{\boldsymbol{i},b}$ be IID random variables with mean 1 variance τ^2 and kurtosis $\kappa_w < \infty$. Then holding $Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} - \bar{X}$ fixed,*

$$\mathrm{Var}_{\mathrm{NBB}}(\widehat{\widetilde{\mathrm{Var}}}_{\mathrm{NBB}}(\bar{X}^*)) = \frac{\sigma^4\tau^4}{BN^2}\left(2 + \frac{\kappa(\kappa_x + 3)}{N}\right)$$

where $\sigma^2 = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} Y_{\boldsymbol{i}}^2$, and $\kappa_x = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} Y_{\boldsymbol{i}}^4/\sigma^4 - 3$. A delta method approximation gives

$$\mathrm{Var}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}}) \doteq \frac{\sigma^4\tau^4}{BN^2}\left(\frac{2B}{B-1} + \frac{\kappa(\kappa_x + 3)}{N}\right).$$

Lee and Clyde (2004), following Oza (2001) view the Poisson version as a lossy online approximation to the bootstrap. Lee and Clyde (2004) prefer exponential weights because it is an exact online version of the Bayesian bootstrap. We find here that there are only small differences between weighting schemes, but double-or-nothing weights having the smallest possible kurtosis $\kappa = -2$ have the best stability. The Poi(1) distribution has $\kappa = 1$ and the Exp(1) distribution has $\kappa = 6$. When $\kappa(\kappa_x + 3) \ll N$, then $\widetilde{\widetilde{\mathrm{Var}}}_{\mathrm{NBB}}(\bar{X}^*)$ with B reweightings has approximately the variance of $s^2_{\mathrm{NBB}}(\bar{X}^*)$ with $B+1$ reweightings.

4 Factorial reweighting

Our proposal here is to apply a product of independent random weights to the data. Observation $\boldsymbol{i}$ is given weight $W_{\boldsymbol{i}} \geq 0$. The weights take the form

$$W_{\boldsymbol{i}} = \prod_{j=1}^{r} W_{j,i_j} \tag{11}$$

where W_{j,i_j} are independent random variables for $j \in [r]$ and $i_j \geq 1$. We assume that $\mathbb{E}(W_{j,i_j}) = 1$ and $\mathrm{Var}(W_{j,i_j}) = \tau_j^2 < \infty$. The usual choice has all τ_j^2 equal to a common τ^2 which in turn is usually equal to 1.

The reweighted mean $\bar{X}^*$ is once again a ratio estimate with delta method approximation

$$\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*) = \frac{1}{N^2}\mathbb{E}_{\mathrm{PW}}\big(\big(T^* - \bar{X}N^*\big)^2\big) \tag{12}$$

where $T^* = \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}} X_{\boldsymbol{i}}$ and $N^* = \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}}$ for $W_{\boldsymbol{i}}$ given by (11). The subscript PW refers to random weights taking the product form.

The bootstrap variance depends on precise details of the overlaps among different observations. We will derive some approximations to this variance below. For the exact variance we need to introduce some additional quantities:

$$\begin{aligned}
\rho_k &= \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k},\\
\nu_{k,u} &= \frac{1}{N}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}\mathbf{1}_{i_u=i'_u}, \quad \text{and,}\\
\widetilde{\nu}_{k,u} &= \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}\sum_{\boldsymbol{i}''} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} Z_{\boldsymbol{i}''} \mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}\mathbf{1}_{i_u=i'''_u}\\
&= \frac{1}{N^2}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} N_{\boldsymbol{i},u} N_{\boldsymbol{i},k}
\end{aligned}$$

for $k = 0, 1, \ldots, r$ and $u \subseteq [r]$. In words, ρ_k gives the fraction of data pairs that match in exactly k positions, while $\nu_{k,u}/N$ gives the fraction of data pairs that match in exactly k positions including all $j \in u$. The third quantity, $\widetilde{\nu}_{k,u}$, is N

times the fraction of data triples $(\boldsymbol{i}, \boldsymbol{i}', \boldsymbol{i}'')$ in which $\boldsymbol{i}$ matches $\boldsymbol{i}'$ in precisely k places while also matching $\boldsymbol{i}''$ for all $j \in u$.

These new quantities satisfy the identities

$$\sum_{k=0}^{r} \rho_k = 1, \quad \text{and} \quad \sum_{k=0}^{r} \nu_{k,u} = \sum_{k=0}^{r} \widetilde{\nu}_{k,u} = \nu_u.$$

Also it is clear that $\nu_{k,u} = 0$ when $|u| > k$.

Theorem 5. *In the random effects model* (1)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N}\sum_{u \neq \varnothing} \gamma_u \sigma_u^2, \tag{13}$$

where

$$\gamma_u = \sum_{k=0}^{r} (1+\tau^2)^k (\nu_{k,u} - 2\widetilde{\nu}_{k,u} + \rho_k \nu_u). \tag{14}$$

The quantities γ_u are 'gain coefficients' which multiply σ_u^2/N. Ideally they should equal ν_u and then the bootstrap variance would match the desired one. Where they differ from ν_u, the bootstrap variance is biased. Typically the bias is positive, making this bootstrap conservative. Sometimes the bias is very small.

The special case $r = 1$ is interesting because it corresponds to IID sampling. Then the only variance component is $\sigma^2_{\{1\}}$ which we abbreviate to σ^2 and equation (13) simplifies to

$$\frac{\tau^2\sigma^2}{N}\Big(1 - \frac{1}{N}\Big) = \frac{\tau^2\sigma^2}{N-1}.$$

In this instance there is a (trivial) negative bias if $\tau^2 = 1$.

Independently reweighting rows and columns is similar to independently resampling them. That strategy of bootstrapping rows and columns has been given several names in the literature. Brennan et al. (1987) called it "boot-p,i" because for educational testing data, it resamples both people and items. McCullagh (2000, page 294) calls the method "Boot-II". There is also another "Boot-II" for the one way layout in that paper. Noting a similarity to Cornfield and Tukey's pigeonhole model for analysis of variance, Owen (2007) calls this approach the "pigeonhole bootstrap". Reweighting with a product of Rubin's (1981) exponential weights is thus a "Bayesian pigeonhole bootstrap".

5 Interpretable approximations

Theorem 5 gives exact finite sample formulas for the gain coefficients γ_u, but they are unwieldy. Here we make some approximations to γ_u in order to get more interpretable results.

First we introduce the quantity

$$\epsilon = \max_{\boldsymbol{i}} \max_{u \neq \varnothing} \frac{N_{\boldsymbol{i},u}}{N} = \max_{\boldsymbol{i}} \max_{1 \le j \le r} \frac{N_{\boldsymbol{i},\{j\}}}{N},$$

which measures the largest proportional duplication of indices. Though $1 \ge \epsilon \ge 1/N$, we anticipate that ϵ will usually be small. For the Netflix data, $\epsilon = 232{,}944/100{,}480{,}507 \doteq 0.00232$ stemming from one movie having 232,944 ratings.

Although we suppose that ϵ is small below, it is worth pointing out that exceptions do arise, even for some very large data sets. For example if the observed data form a complete $N_1 \times N_2 \times \cdots \times N_r$ sample, then $\epsilon = \max_{1 \le j \le r} 1/N_j$. If one factor takes only a modest number of levels, then ϵ is large. A second context where ϵ is large arises when one of the factors is greatly dominated by one of its levels, as for example, we might find in internet data where one factor is the country of the web user.

A second parameter to aid interpretability is

$$\eta = \max_{\varnothing \subsetneq u \subsetneq v} \frac{\nu_v}{\nu_u}.$$

By construction $\eta \le 1$, and we ordinarily expect η to be small. Of the indices which match for $j \in u$ only a relatively small number should also match for $j \in v - u$ too, because each additional match in large data sets represents a coincidence. For the Netflix data

$$\eta = \max\{\nu_{\{1,2\}}/\nu_{\{1\}}, \nu_{\{1,2\}}/\nu_{\{2\}}\} = 1/646 \doteq 0.00155.$$

While η is often small, there are exceptions. If two factors are very dependent then η need not be small. For example people's names and phone numbers may be such variables: many or even most phone numbers are used by a small number of people (often one) and many people use only a small number of phone numbers. Then the fraction of data pairs matching on both of these variables will not be much smaller than the fraction matching on one of them.

In simplifying expressions we use $O(\eta)$ and $O(\epsilon)$. These describe limits as η (respectively ϵ) converge to 0. The implied constants may depend on r. In some expressions we have retained explicit constants.

Theorem 6. *In the random effects model* (1)*, the gain coefficient* (14) *for* $u \neq \varnothing$ *in the product reweighted bootstrap is*

$$\gamma_u = \nu_u[(1+\tau^2)^{|u|} - 1 + \Theta_u \varepsilon] + \sum_{v \supsetneq u} (1+\tau^2)^{|v|} (\tau^2)^{|v-u|} \nu_v \tag{15}$$

where $|\Theta_u| \le (1+\tau^2)((1+\tau^2)^r - 1)/\tau^2$*. For* $\tau^2 = 1$*,*

$$\gamma_u = \nu_u[2^{|u|} - 1 + \Theta_u \varepsilon] + \sum_{v \supsetneq u} 2^{|v|} \nu_v,$$

where $|\Theta_u| \le 2^{r+1} - 2$*.*

For $r = 2$ using $\nu_{\{1,2\}} = 1$ and the usual choice $\tau^2 = 1$, we find that

$$\begin{aligned}\gamma_{\{j\}} &= \nu_{\{j\}}(1 + \Theta_{\{j\}}\epsilon) + 2, \quad j = 1, 2, \quad \text{and}\\ \gamma_{\{1,2\}} &= \nu_{\{1,2\}}(3 + \Theta_{\{1,2\}}\epsilon)\end{aligned}$$

where each $|\Theta| \le 6$. The Bayesian pigeonhole bootstrap variance closely matches the ordinary pigeonhole bootstrap variance. In the extreme setting where $\sigma^2_{\{1\}} = \sigma^2_{\{2\}} = 0 < \sigma^2_{\{1,2\}}$ the resulting bootstrap variance is about three times as high as it should be. In a limit as $\min_j \nu_{\{j\}} \to \infty$ and $\epsilon \to 0$,

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} \to 1 \tag{16}$$

holds for fixed $\sigma^2_{\{j\}} > 0$, $j = 1, 2$. For $r = 3$, with $\nu_{\{1,2,3\}} = 1$ and $\tau^2 = 1$

$$\begin{aligned}\gamma_{\{1\}} &\approx \nu_{\{1\}} + 4\nu_{\{1,2\}} + 4\nu_{\{1,3\}} + 8\\ \gamma_{\{1,2\}} &\approx 3\nu_{\{1,2\}} + 8, \quad \text{and}\\ \gamma_{\{1,2,3\}} &\approx 7,\end{aligned}$$

where $\approx$ reflects an additive error of size $\nu_u\Theta_u\varepsilon$ for $|\Theta_u| \le 14$. In the extreme case where the only nonzero variance coefficient is $\sigma^2_{[3]}$ then the product reweighted bootstrap variance is about 7 times as large as it should be. On the other hand, when the main effect variances $\sigma^2_{\{j\}}$ are positive and $\nu_v/\nu_u \to 0$ for $v \subsetneq u$, then (16) holds. More generally, we have Theorem 7.

Theorem 7. *For the random effects model* (1) *and the product reweighted bootstrap with* $\tau^2 = 1$*, the gain coefficient for nonempty* $u \subseteq [r]$ *satisfies*

$$2^{|u|} - 1 - (2^{r+1} - 2)\epsilon < \frac{\gamma_u}{\nu_u} \le 2^{|u|}(1 + 2\eta)^{|v-u|} - 1 + (2^{r+1} - 2)\epsilon.$$

If there exist m *and* M *with* $0 < m \le \sigma^2_u \le M < \infty$ *for all* $u \ne \varnothing$*, then*

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = 1 + O(\eta + \epsilon).$$

The first claim of Theorem 7 can be summarized as

$$\frac{\gamma_u}{\nu_u} = (2^{|u|} - 1)(1 + O(\eta)) + O(\epsilon) \approx 2^{|u|} - 1,$$

and the second as $\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)/\mathrm{Var}(\bar{X}) \approx 1$, where the implied constants on r. They generally grow exponentially in r but the interesting values of r are small integers from 2 to 6 or so. The main effects dominate when η is small and they are properly accounted for when ϵ is small.

6 The heteroscedastic model

In the r-fold crossed random effects model (1), the term $\varepsilon_{\boldsymbol{i},u}$ has the same variance for all $\boldsymbol{i}$. This model may not be realistic. For instance, the Netflix data includes some customers whose ratings have very small variance and others with a very large variance. Similarly, but to a lesser extent, movies also differ in the variance of their ratings. Unequal variances have the potential to bias inferences, especially in unbalanced cases, because the entities with more observations on them might have systematically higher variance than the others do.

A more realistic model is the ***heteroscedastic r-fold crossed random effects model***, with

$$X_{\boldsymbol{i}} = \mu + \sum_{u \neq \varnothing} \varepsilon_{\boldsymbol{i},u} \tag{17}$$

where $\mu \in \mathbb{R}$ and $\varepsilon_{\boldsymbol{i},u}$ are independent random variables with mean 0 and variance $\sigma^2_{\boldsymbol{i},u}$. There are more variance parameters than observations, we do not need to estimate them. Owen (2007) gives conditions under which the pigeonhole bootstrap with $r = 2$ produces a variance estimate with relative error tending to zero in the heteroscedastic setting. Here we investigate product reweighting with general r for model (17).

We need some new quantities. For $u \neq \varnothing$, define

$$\begin{aligned}
\nu_{\boldsymbol{i},u} &= \frac{1}{N}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'}\mathbf{1}_{i_u = i'_u} = \frac{N_{\boldsymbol{i},u}}{N},\\
\nu_{\boldsymbol{i},k} &= \frac{1}{N}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'}\mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k} = \frac{N_{\boldsymbol{i},k}}{N}, \quad \text{and}\\
\nu_{\boldsymbol{i},k,u} &= \frac{1}{N}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'}\mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}\mathbf{1}_{i_u = i'_u}.
\end{aligned}$$

We also will use

$$\begin{aligned}
\overline{\nu_u\sigma^2_u} &= \frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u}, \quad \text{and}\\
\overline{\nu_k} &= \frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\nu_{\boldsymbol{i},k}.
\end{aligned}$$

Next, we parallel the development from the ordinary random effects model (1). Theorem 8 gives the exact variance of $\bar{X}$ for heteroscedastic random effects, Theorem 9 gives the gain coefficients under product reweighting, Theorem 10 provides interpretable bounds for the gains in terms of ϵ. Finally Theorem 11 gives conditions under which the product reweighted bootstrap has a negligible bias.

Theorem 8. *In the heteroscedastic random effect model* (17)

$$\mathrm{Var}(\bar{X}) = \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}} \nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u}. \tag{18}$$

Theorem 9. *In the heteroscedastic random effects model* (17)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u}, \tag{19}$$

where

$$\gamma_{\boldsymbol{i},u} = \sum_{k=0}^{r}(1+\tau^2)^k(\nu_{\boldsymbol{i},k,u} - 2\nu_{\boldsymbol{i},k}\nu_{\boldsymbol{i},u} + \overline{\nu_k}\nu_{\boldsymbol{i},u}). \tag{20}$$

Theorem 10. *In the heteroscedastic random effects model* (17)*, the gain coefficient* $\gamma_{\boldsymbol{i},u}$ *of* (20) *for* $Z_{\boldsymbol{i}} = 1$ *and* $u \neq \varnothing$*, in the product reweighted bootstrap is*

$$\gamma_{\boldsymbol{i},u} = \nu_{\boldsymbol{i},u}[(1+\tau^2)^{|u|} - 1 + \Theta_u\varepsilon] + \sum_{v\supsetneq u}(1+\tau^2)^{|v|}(\tau^2)^{|v-u|}\nu_{\boldsymbol{i},v}$$

where $|\Theta_u| \le (1+\tau^2)((1+\tau^2)^r - 1)/\tau^2$*. For* $\tau^2 = 1$

$$\gamma_{\boldsymbol{i},u} = \nu_{\boldsymbol{i},u}[2^{|u|} - 1 + \Theta_u\varepsilon] + \sum_{v\supsetneq u}2^{|v|}\nu_{\boldsymbol{i},v}$$

where $|\Theta_u| \le 2^{r+1} - 2$*.*

Theorem 10 establishes that our bootstrap is conservative in the heteroscedastic case. With $\tau^2 = 1$ we have

$$\frac{\gamma_{\boldsymbol{i},u}}{\nu_{\boldsymbol{i},u}} \ge 2^{|u|} - 1 - (2^{r+1} - 2)\epsilon.$$

For the homoscedastic random effects model, the main effects dominate when $\eta = \max_{\varnothing\subsetneq u\subsetneq v}\nu_v/\nu_u$ is small and the variance components are all within the interval $[m, M]$ for $0 < m \le M > \infty$. In the heteroscedastic case we might reasonably require every $\sigma^2_{\boldsymbol{i},u} \in [m, M]$. The analysis we used for Theorem 7 also requires the quantities

$$\eta_{\boldsymbol{i}} = \begin{cases} \max_{\varnothing\subsetneq u\subsetneq v}\dfrac{\nu_{\boldsymbol{i},v}}{\nu_{\boldsymbol{i},u}} & Z_{\boldsymbol{i}} = 1 \\ 0 & Z_{\boldsymbol{i}} = 0 \end{cases}$$

to be small.

For $r = 2$ the only subsets u and v which appear in $\eta_{\boldsymbol{i}}$ are $u = \{j\}$ and $v = \{1, 2\}$. Furthermore $\nu_{\boldsymbol{i},\{1,2\}} = 1/N$ and so

$$\max_{\boldsymbol{i}}\eta_{\boldsymbol{i}} = \max_{j\in\{1,2\}}\max_{\boldsymbol{i}}\frac{N_{\boldsymbol{i},\{j\}}}{N} = \epsilon.$$

Then using the same argument we used to prove the second part of Theorem 7 we get

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = 1 + O(\epsilon), \qquad \text{for } r = 2.$$

The case for $r > 2$ is more complicated. There may be observations $\boldsymbol{i}$ with large values for $\nu_{\boldsymbol{i},v}/\nu_{\boldsymbol{i},u}$ where $\varnothing \subsetneq u \subsetneq v$. We still get a good approximation from the product reweighted bootstrap because even though the individual $\eta_{\boldsymbol{i}}$ need not always be small, sums of $\nu_{\boldsymbol{i},v}$ over i are small compared to corresponding sums of $\nu_{\boldsymbol{i},u}$ for $\varnothing \subsetneq u \subsetneq v$.

Theorem 11. *For the heteroscedastic random effects model* (17)*, assume that there exist* m *and* M *with* $0 < m \le \sigma^2_{\boldsymbol{i},u} \le M < \infty$*. Then the product reweighted bootstrap with* $\tau^2 = 1$*, satisfies*

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = 1 + O(\eta + \epsilon).$$

7 Nested random effects

The r-fold crossed random effects model (1) excludes replicated observations by definition: there can be only one $X_{\boldsymbol{i}}$ for any combination $\boldsymbol{i}$ of factors. If two X's are observed to share all index values i_j, we can incorporate them by introducing an $r+1$'st index i_{r+1} which breaks the ties. Conditionally on the effects of the first r indices, distinct replicates are independent. That is $\sigma^2_u = 0$ when $r+1 \in u$ but $u \neq \{1, 2, \ldots, r+1\}$. The replicate index i_{r+1} is a factor that is nested within the first r factors.

More generally, we could have s additional indices corresponding to factors crossed with each other, but nested within our r outer factors. Then the index $\boldsymbol{i} \in \{1, 2, \ldots\}^{r+s}$ uniquely identifies a data point. Ordinary replication has $s = 1$. The nesting structure means that

$$\sigma^2_u = 0, \quad \text{if} \quad u \cap \{r+1, \ldots, r+s\} \neq \varnothing \quad \text{and} \quad u \cap [r] \neq [r]. \tag{21}$$

In words, the effect $\epsilon_{\boldsymbol{i},u}$ is 0 if the factors in u include ***any*** of the inner factors without including ***all*** of the outer factors.

When one factor is nested within another, such as replicates within subjects, it is a common practice to resample or reweight the outer factor only. For example, the resampled data set might contain resampled subjects retaining the repeated measurements from each of them.

In the nested setting, the variance of $\bar{X}$ under and $r+s$ factor version of the random effects model (1) is still $(1/N)\sum_{u \neq \varnothing} \nu_u \sigma^2_u$ although many of the σ^2_u terms are zero.

For $\boldsymbol{i} \in [r+s]$ let $\lfloor \boldsymbol{i} \rfloor = (i_1, \ldots, i_r)$ be the indices of its inner factors. We can study these nested models by introducing the variables

$$T_{\lfloor \boldsymbol{i} \rfloor} = \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'} 1_{\lfloor \boldsymbol{i}' \rfloor = \lfloor \boldsymbol{i} \rfloor} X_{\boldsymbol{i}'}, \quad \text{and}$$
$$M_{\lfloor \boldsymbol{i} \rfloor} = \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}'} 1_{\lfloor \boldsymbol{i}' \rfloor = \lfloor \boldsymbol{i} \rfloor},$$

so that the sample mean

$$\bar{X} = \frac{1}{N}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} X_{\boldsymbol{i}} = \frac{\sum_{\lfloor \boldsymbol{i} \rfloor} T_{\lfloor \boldsymbol{i} \rfloor}}{\sum_{\lfloor \boldsymbol{i} \rfloor} M_{\lfloor \boldsymbol{i} \rfloor}}$$

is an r-factor ratio estimator.

When the numbers $M_{\lfloor \boldsymbol{i} \rfloor}$ of replicates for each outer factor vary, we obtain a heteroscedastic random effects model in the first r variables.

8 Extensions

We have used the variance of a sample mean as a way to identify a suitable bootstrap method. A bootstrap method that under-estimates the variance of a mean cannot be expected to work well on other problems. One that is properly calibrated or conservative for the variance of a scalar sample mean will also work in some other settings.

The bootstrap is usually used to get confidence intervals, not variance estimates. For an asymptotically unbiased statistic that satisfies a central limit theorem, a properly calibrated variance yields asymptotically correct bootstrap percentile confidence intervals. An overestimated variance yields conservative percentile intervals.

When $X_{\boldsymbol{i}} \in \mathbb{R}^d$ for $d > 1$ we may simply replace the variances σ_u^2 or $\sigma_{\boldsymbol{i},u}^2$ by variance-covariance matrices Σ_u or $\Sigma_{\boldsymbol{i},u}$ respectively in the variance formulas. This follows by considering the variance of $\phi^{\mathsf{T}} X_{\boldsymbol{i}}$ for vectors $\phi \in \mathbb{R}^d$.

Bootstrap correctness extends from means to other statistics. See Hall (1992) and Mammen (1992). The extension to smooth functions $g(\bar{X})$ of means is via Taylor expansion, when g has a Jacobian matrix with full rank at $\mathbb{E}(\bar{X})$.

Similarly, the extension from means to maximum likelihood estimates follows by considering estimating equations. We leave out regularity conditions, but note that for a model relating $Y_{\boldsymbol{i}}$ to $X_{\boldsymbol{i}}$ via $(X_{\boldsymbol{i}}, Y_{\boldsymbol{i}}) \sim f(X_{\boldsymbol{i}}, Y_{\boldsymbol{i}}; \beta)$ with parameter $\beta \in \mathbb{R}^p$, we could test $\beta = \beta_0$ by testing whether

$$\frac{\partial f(X_{\boldsymbol{i}}, Y_{\boldsymbol{i}}; \beta)/\partial \beta \mid_{\beta=\beta_0}}{f(X_{\boldsymbol{i}}, Y_{\boldsymbol{i}}; \beta_0)}$$

has mean 0. In practice it is usually more convenient to form a histogram of resampled $\hat{\beta}^*$ values.

We have worked conditionally on the observed values holding $Z_{\boldsymbol{i}}$ fixed. Missingness can introduce a bias and one might seek to adjust for it. Any model for those $X_{\boldsymbol{i}}$ with $Z_{\boldsymbol{i}} = 0$ requires assumptions that cannot be tested within the data set. If we can model the missingness and estimate parameters based on the data at hand by a parametric model, then we can use the bootstrap to judge the sampling variance of our parameter estimates. The bootstrap will not correct for any bias resulting from an incorrect model of missingness.

9 Example: loquacity of Facebook comments

We present an analysis of national differences in comment length on Facebook. In particular, Facebook users can share links with their friends. Their friends, and the posting user, can comment on the link. We compare the length of these comments produced by users in the United States using the site in American English (*US users*) and those produced by users in the United Kingdom using the site in British English (*UK users*). We restrict the analysis to US and UK users commenting on links shared by US and UK users. We additionally consider two different modes by which users can comment: the standard web interface to Facebook (*web*) and an application for some touchscreen mobile phones (*mobile*).

We treat the logarithm of the number of characters in a comment as the outcome in the following random effects model:

$$X_{cm\boldsymbol{i}} = \mu_{cm} + \sum_{u \neq \varnothing} \varepsilon_{\boldsymbol{i},u}$$

where μ_{cm} is the mean log characters for country c in mode m. Here the members of $\boldsymbol{i}$ are indexes for the user sharing the link (*sharer*), the user commenting on the link (*commenter*), and the canonicalized URL being shared (*URL*). By definition, no comments have 0 characters, and so each X in our data set is well defined.

The data consist of $X_{cm\boldsymbol{i}}$ for a sample of comments by US and UK who are using Facebook in American and British English, respectively, during a short period in 2011. This sample includes 18,134,419 comments by 8,078,531 commenters on 2,085,639 URLs shared by 3,904,715 sharers. We examine whether these US and UK users post comments of different lengths for both of the modes. The duplication coefficients for this data are

$$\begin{aligned} \nu_{\text{sh}} &\doteq 17.71, \quad \nu_{\text{com}} \doteq 7.71, \quad \nu_{\text{url}} \doteq 26{,}854.92 \\ \nu_{\text{sh,com}} &\doteq 5.92, \quad \nu_{\text{sh,url}} \doteq 12.91, \quad \nu_{\text{com,url}} \doteq 5.19, \quad \text{and} \\ \nu_{\text{sh,com,url}} &\doteq 4.88. \end{aligned}$$

The coefficient for URLs is conspicuously large, indicating that a naive bootstrap would be very unreliable.

The sample mean for a country and mode is

$$\hat{\mu}_{cm} = \frac{\sum_{\boldsymbol{i}} Z_{cm\boldsymbol{i}} X_{cm\boldsymbol{i}}}{\sum_{\boldsymbol{i}} Z_{cm\boldsymbol{i}}}.$$

We regard $\hat{\mu}_{cm}$ as an estimate of μ_{cm} conditional on the observed combinations of sharers, commenters, and URLs.

The four sample means for both countries and both modes suggest that the US users write longer comments than UK users when commenting on the web ($\hat{\mu}_{\text{US, web}} = 3.62$, $\hat{\mu}_{\text{UK, web}} = 3.55$), while UK users write longer comments than US users when commenting via the selected mobile interface ($\hat{\mu}_{\text{US, mobile}} =$

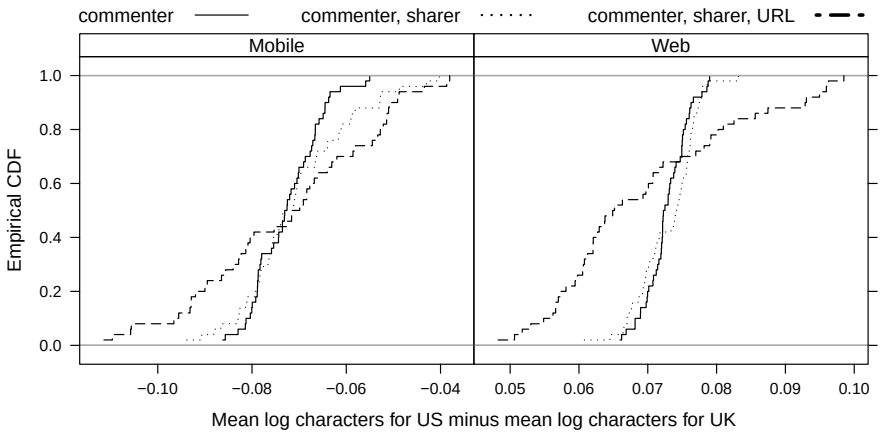

Figure 1: Difference between the logged number of characters in comments by US and UK users for three different bootstrap reweightings with $R = 50$. Each data point in the plotted ECDF is the difference in means from a single bootstrap reweighting. US users post longer comments than UK users on the web, but this difference is reversed for the mobile interface studied.

3.5, $\hat{\mu}_{\text{UK, mobile}} = 3.57$). Many differences between US and UK users likely contribute to this observed differences. Before searching for causes of these two differences, a data analyst would likely want to quantify the evidence for the existence and size of these differences. We test whether these two pairs of means are likely to be observed given the null hypothesis of no difference in comment length between the countries within each mode.

Using software for Hive (Thusoo et al., 2009), a Hadoop-based map–reduce data warehousing and parallel computing environment, we can compute each of these four means for a number of bootstrap reweightings of the data, while visiting each observation only once. When visiting an observation, the hashed identifiers for the factor levels for that observation are each used as seeds to random number generators. This allows all nodes to use the same $U\{0, 2\}$ draw in computing the product weight for all observations that share a particular factor level. Note that users can be both sharers and commenters. Since users can comment on their own shared links, some observations could have the same factor level identifier for both the sharer and commenter levels. We use different portions of the hashed identifier so that the weights for these two roles are not dependent. For each reweighting, we compute four reweighted sample means

$$\hat{\mu}^*_{cm} = \frac{\sum_{\boldsymbol{i}} Z_{cm\boldsymbol{i}} W_{cm\boldsymbol{i}} X_{cm\boldsymbol{i}}}{\sum_{\boldsymbol{i}} Z_{cm\boldsymbol{i}} W_{cm\boldsymbol{i}}},$$

corresponding to $c \in \{\text{US}, \text{UK}\}$ and $m \in \{\text{web}, \text{mobile}\}$.

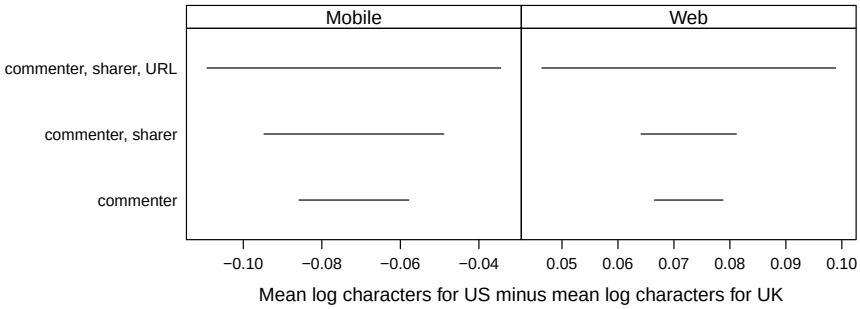

Figure 2: Confidence intervals for the difference between the logged number of characters in comments by US and UK users for three different bootstrap reweightings with $R = 50$. Confidence intervals span the 2.5% and 97.5% quantiles of the normal with variance computed from the bootstrap reweightings. While all three analyses reject the null hypothesis, the one- and two-factor analyses may substantially overstate confidence about the size of the true difference, especially in the case of comments posted via the web interface.

For comparison, we conduct this analysis reweighting one, two, and all three of the factors. Figure 1 presents $R = 50$ bootstrapped differences in the two pairs of means when reweighting commenters, commenters and sharers, and all three factors. Inspection of these ECDFs confirms that the observed differences cannot be attributed to chance, even when accounting for the random main and interaction effects of commenters, sharers, and URLs. The bootstrapped differences in means are strikingly more dispersed for the three-factor analysis. Figure 2 shows 95% confidence intervals for the two differences computed as quantiles of the normal distribution with variance computed from the bootstrap reweightings. This highlights the substantial overstatement of certainty that can come from neglecting the presence of additional random effects. In this case, the three analyses would all reject the null hypothesis, but would produce quite different confidence intervals.

For the approximations developed in Section 5 to apply, we require that ϵ and η be small – that no single level of any random effect make up a large portion of the observations and that the number of observations matching on v is small compared to the number matching on u factors for all $\varnothing \subsetneq u \subsetneq v$. We find that $\epsilon = 686{,}990/18{,}134{,}419 \doteq 0.0379$, as one URL had 686,990 comments in this sample. We also found that $\eta \doteq 0.767$. Because η is not very small it is possible that the variance estimates are conservative.

Acknowledgments

This work was supported by grant DMS-0906056 from the U.S. National Science Foundation and by Nokia. We thank Paul Jones and Jonathan Chang for their assistance.

References

Bennett, J. and Lanning, S. (2007). The Netflix prize. In *Proceedings of KDD Cup and Workshop 2007*.

Brennan, R. L., Harris, D. J., and Hanson, B. A. (1987). The bootstrap and other procedures for examining the variability of estimated variance components. Technical report, ACT.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Lee, H. K. H. and Clyde, M. A. (2004). Lossless online Bayesian bagging. *Journal of Machine Learning Research*, 5:143–151.

Mammen, E. (1992). *When Does Bootstrap Work?*, volume 77 of *Lecture Notes in Statistics*. Springer, New York.

McCarthy, P. J. (1969). Pseudo-replication: half samples. *Review of the International Statistical Institute*, 37(3):239–264.

McCullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli*, 6(2):285–301.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (disc: P26-48). *Journal of the Royal Statistical Society, Series B, Methodological*, 56:3–26.

Owen, A. B. (2007). The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2):386–411.

Oza, N. (2001). Online bagging and boosting. In *Systems, man and cybernetics, 2005 IEEE international conference on*, volume 3, pages 2340–2345. IEEE.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9:130–134.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York.

Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Anthony, S., Liu, H., Wyckoff, P., and Murthy, R. (2009). Hive: a warehousing solution over a map-reduce framework. In *Proceedings of the VLDB Endowment*, volume 2, pages 1626–1629. VLDB Endowment.

Wiley, E. W. (2001). *Bootstrap strategies for variance component estimation: theoretical and empirical results*. PhD thesis, Stanford University.

Appendix: proofs

This appendix contains theorem proofs and a few lemmas. The theorems are restated to make it easier to follow the steps. Equation numbers that appear in the theorem statements from the article are preserved in this appendix.

Proof of Theorem 1

Theorem 1. *In the random effects model* (1)

$$\mathrm{Var}(\bar{X}) = \frac{1}{N}\sum_{u\neq\varnothing}\nu_u\sigma_u^2.$$

Proof. The numerator of $\bar{X}$ in (3) is $\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}X_{\boldsymbol{i}} = N\mu + \sum_{\boldsymbol{i}}\sum_{u\neq\varnothing} Z_{\boldsymbol{i}}\varepsilon_{\boldsymbol{i}_u}$. Therefore the variance of $\bar{X}$ under the random effects model is

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \frac{1}{N^2}\mathbb{E}\Big(\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\sum_{u\neq\varnothing}\sum_{u'\neq\varnothing}\varepsilon_{\boldsymbol{i},u}\varepsilon_{\boldsymbol{i}',u'}\Big)\\
&= \frac{1}{N^2}\sum_{u\neq\varnothing}\sigma_u^2\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbf{1}_{i_u=i'_u}\\
&= \frac{1}{N^2}\sum_{u\neq\varnothing}\sigma_u^2\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}N_{\boldsymbol{i},u}\\
&= \frac{1}{N}\sum_{u\neq\varnothing}\nu_u\sigma_u^2.
\end{aligned}$$

□

Proof of Theorems 2, 3, and 4.

Here we prove the theorems about naive bootstrap sampling. Theorem 2 is about naive resampling and Theorem 3 handles naive reweighting. Theorem 4 is about bootstrap stability.

Theorem 2. *Under the random effects model* (1)*, the expected value of the naive bootstrap variance of* $\bar{X}$ *is*

$$\mathbb{E}_{\mathrm{RE}}(\mathrm{Var}_{\mathrm{NB}}(\bar{X})) = \frac{1}{N}\sum_{u\neq\varnothing}\sigma_u^2\Big(1-\frac{\nu_u}{N}\Big). \tag{6}$$

Proof. A U-statistic decomposition of the sample variance is

$$\begin{aligned}\mathrm{Var}_{\mathrm{NB}}(\bar{X}) &= \frac{1}{2N^3}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}(X_{\boldsymbol{i}}-X_{\boldsymbol{i}'})^2 \\ &= \frac{1}{2N^3}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\Big(\sum_{u\neq\varnothing}\varepsilon_{\boldsymbol{i},u}-\varepsilon_{\boldsymbol{i}',u}\Big)^2.\end{aligned}$$

Under the random effects model

$$\begin{aligned}\mathbb{E}_{\mathrm{RE}}\big(\mathrm{Var}_{\mathrm{NB}}(\bar{X})\big) &= \frac{1}{2N^3}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\sum_{u\neq\varnothing}2\sigma_u^2(1-\mathbf{1}_{i_u=i'_u}) \\ &= \frac{1}{N}\sum_{u\neq\varnothing}\sigma_u^2\Big(1-\frac{\nu_u}{N}\Big).\end{aligned}$$

□

To prove Theorem 3, we begin with a lemma on the covariance of pairs of observations under the random effects model.

Lemma 1. *Let $X_{\boldsymbol{i}}$ follow the random effects model* (1) *and let $Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} - \bar{X}$. Then*

$$\mathbb{E}_{\mathrm{RE}}(X_{\boldsymbol{i}}X_{\boldsymbol{i}'}) = \mu^2 + \sum_{u\neq\varnothing}\sigma_u^2\mathbf{1}_{i_u=i'_u} \tag{22}$$

and

$$\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}) = \sum_{u\neq\varnothing}\sigma_u^2\Big(\mathbf{1}_{i_u=i'_u} - \frac{N_{\boldsymbol{i},u}}{N} - \frac{N_{\boldsymbol{i}',u}}{N} + \frac{\nu_u}{N}\Big). \tag{23}$$

Proof. Equation (22) follows directly from the random effects model definition. Expanding $Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}$ yields

$$X_{\boldsymbol{i}}X_{\boldsymbol{i}'} - \frac{1}{N}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}X_{\boldsymbol{i}}X_{\boldsymbol{i}''} - \frac{1}{N}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}X_{\boldsymbol{i}'}X_{\boldsymbol{i}''} + \frac{1}{N^2}\sum_{\boldsymbol{i}''}\sum_{\boldsymbol{i}'''}Z_{\boldsymbol{i}''}Z_{\boldsymbol{i}'''}X_{\boldsymbol{i}''}X_{\boldsymbol{i}'''}.$$

Because μ cancels from $Y_{\boldsymbol{i}}$ we may assume that $\mu = 0$ while proving (23). Now

$$\mathbb{E}_{\mathrm{RE}}\Big(\frac{1}{N}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}'}X_{\boldsymbol{i}}X_{\boldsymbol{i}'}\Big) = \frac{1}{N}\sum_{u\neq\varnothing}\sigma_u^2\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}'}\mathbf{1}_{i_u=i'_u} = \frac{1}{N}\sum_{u\neq\varnothing}\sigma_u^2 N_{\boldsymbol{i},u}.$$

Therefore

$$\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}) = \sum_{u\neq\varnothing}\sigma_u^2\Big(\mathbf{1}_{i_u=i'_u} - \frac{N_{\boldsymbol{i},u}}{N} - \frac{N_{\boldsymbol{i}',u}}{N} + \frac{1}{N^2}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}N_{\boldsymbol{i}'',u}\Big)$$

which reduces to (23). □

Theorem 3. *In the random effects model* (1)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*)\big) = \frac{\tau^2}{N}\sum_{u\neq\varnothing}\sigma_u^2\Big(1-\frac{\nu_u}{N}\Big). \tag{7}$$

Proof. Let $Y_i = X_i - \bar{X}$ and $T_y^* = \sum_{\boldsymbol{i}} W_{\boldsymbol{i}} Z_{\boldsymbol{i}} Y_{\boldsymbol{i}}$. Then

$$\begin{aligned}
\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*)\big) &= \frac{1}{N^2}\mathbb{E}_{\mathrm{RE}}\big(\mathbb{E}_{\mathrm{NBB}}\big((T^*-\bar{X}N^*)^2\big)\big)\\
&= \frac{1}{N^2}\mathbb{E}_{\mathrm{RE}}\Big(\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}\mathbb{E}_{\mathrm{NBB}}(W_{\boldsymbol{i}}W_{\boldsymbol{i}'})\Big)\\
&= \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'})\mathbb{E}_{\mathrm{NBB}}(W_{\boldsymbol{i}}W_{\boldsymbol{i}'}).
\end{aligned}$$

Next, $\mathbb{E}_{\mathrm{NBB}}(W_{\boldsymbol{i}}W_{\boldsymbol{i}'}) = 1+\tau^2\mathbf{1}_{\boldsymbol{i}=\boldsymbol{i}'}$. Therefore

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*)\big) = \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}) + \frac{\tau^2}{N^2}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}^2). \tag{24}$$

The double sum in (24) vanishes because $\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}Y_{\boldsymbol{i}} = 0$. Then from Lemma 1, the coefficient of σ_u^2 in (24) is

$$\frac{\tau^2}{N^2}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\Big(1-\frac{2N_{\boldsymbol{i},u}}{N}+\frac{\nu_u}{N}\Big) = \frac{\tau^2}{N^2}\big(N-2\nu_u+\nu_u\big)$$

establishing (7). □

Theorem 4. *Let W and $W_{\boldsymbol{i},b}$ be IID random variables with mean 1 variance τ^2 and kurtosis $\kappa_w < \infty$. Then holding $Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} - \bar{X}$ fixed,*

$$\mathrm{Var}_{\mathrm{NBB}}(\widehat{\widetilde{\mathrm{Var}}}_{\mathrm{NBB}}(\bar{X}^*)) = \frac{\sigma^4\tau^4}{BN^2}\left(2+\frac{\kappa(\kappa_x+3)}{N}\right)$$

where $\sigma^2 = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}Y_{\boldsymbol{i}}^2$, and $\kappa_x = (1/N)\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}Y_{\boldsymbol{i}}^4/\sigma^4 - 3$. A delta method approximation gives

$$\mathrm{Var}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}}) \doteq \frac{\sigma^4\tau^4}{BN^2}\left(\frac{2B}{B-1}+\frac{\kappa(\kappa_x+3)}{N}\right).$$

Proof. First, the variance of $\widehat{\widetilde{\mathrm{Var}}}_{\mathrm{NBB}}(\bar{X}^*)$ scales as $1/B$ so we can work with $B=1$ and divide the result by B. For $B=1$, we drop the subscript b from

W's. We will use the identity $\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}} Y_{\boldsymbol{i}} = \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}(W_{\boldsymbol{i}} - 1) Y_{\boldsymbol{i}}$. If $B = 1$, then $\mathrm{Var}_{\mathrm{NBB}}(\widehat{\mathrm{Var}}_{\mathrm{NBB}}(\bar{X}^*))$ equals

$$\begin{aligned}
&\mathbb{E}_{\mathrm{NBB}}\left(\left(\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}} Y_{\boldsymbol{i}}\right)^4\right) - \left(\frac{\sigma^2\tau^2}{N}\right)^2 \\
&= \frac{1}{N^4}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\mathbb{E}((W_{\boldsymbol{i}} - 1)^4) Y_{\boldsymbol{i}}^4 + \frac{3}{N^4}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'}\mathbb{E}((W_{\boldsymbol{i}} - 1)^2)^2 Y_{\boldsymbol{i}}^2 Y_{\boldsymbol{i}'}^2 \\
&\quad - \frac{3}{N^4}\sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}\mathbb{E}((W_{\boldsymbol{i}} - 1)^2)^2 Y_{\boldsymbol{i}}^4 - \left(\frac{\sigma^2\tau^2}{N}\right)^2 \\
&= \frac{\tau^4\sigma^4(\kappa + 3)(\kappa_x + 3)}{N^3} + \frac{3\tau^4\sigma^4}{N^2} - \frac{3\tau^4\sigma^4(\kappa_x + 3)}{N^3} - \frac{\sigma^4\tau^4}{N^2} \\
&= \frac{\tau^4\sigma^4}{N^2}\left(2 + \frac{\kappa(\kappa_x + 3)}{N}\right).
\end{aligned}$$

For the second part

$$\mathrm{Var}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}}) = \mathbb{E}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}})^2\left(\frac{2}{B-1} + \frac{\kappa^*}{B}\right)$$

where κ^* is the kurtosis of $\bar{X}^* = \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}} Y_{\boldsymbol{i}} / \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}} W_{\boldsymbol{i}}$. The delta method approximation to $\mathbb{E}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}})$ is $\tau^2\sigma^2/N$. For the kurtosis, we make the Taylor approximation

$$\bar{X}^* \doteq \bar{X} + \sum_{\boldsymbol{i}} Z_{\boldsymbol{i}}(W_{\boldsymbol{i}} - 1) Y_{\boldsymbol{i}}.$$

The expected value of $\bar{X}^* - \bar{X}$ reuses much of the above computation and yields

$$\mathbb{E}_{\mathrm{NBB}}((\bar{X}^* - \bar{X})^4) \doteq \frac{\tau^4\sigma^4}{N^2}\left(3 + \frac{\kappa(\kappa_x + 3)}{N}\right).$$

Therefore $\kappa^* = \kappa(\kappa_x + 3)/N$ and so

$$\mathrm{Var}_{\mathrm{NBB}}(s^2_{\mathrm{NBB}}) = \frac{\tau^4\sigma^4}{BN^2}\left(\frac{2B}{B-1} + \frac{\kappa(\kappa_x + 3)}{N}\right). \quad \square$$

Proof of Theorems 5, 6, and 7.

Theorem 5 gives an exact expression for the gain coefficients of the Bayesian pigeonhole bootstrap in the constant variance crossed random effects model. Theorem 6 gives an interpretable approximation to those gain coefficients. Theorem 7 shows factorial reweighting gives nearly the correct variance when ϵ and η are both small.

Theorem 5. *In the random effects model* (1)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N}\sum_{u \neq \varnothing} \gamma_u \sigma_u^2, \tag{13}$$

where

$$\gamma_u = \sum_{k=0}^{r} (1+\tau^2)^k (\nu_{k,u} - 2\widetilde{\nu}_{k,u} + \rho_k \nu_u). \tag{14}$$

Proof. We begin along the same lines as Theorem 3 and find that

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N^2} \sum_{\boldsymbol{i}} \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}} Y_{\boldsymbol{i}'}) \mathbb{E}_{\mathrm{PW}}(W_{\boldsymbol{i}} W_{\boldsymbol{i}'}).$$

For the product weights used in this bootstrap,

$$\mathbb{E}_{\mathrm{PW}}(W_{\boldsymbol{i}} W_{\boldsymbol{i}'}) = \prod_{j : i_j = i'_j} (1+\tau^2) = (1+\tau^2)^{|M_{\boldsymbol{i}\boldsymbol{i}'}|},$$

with $\mathbb{E}_{\mathrm{PW}}(W_{\boldsymbol{i}} W_{\boldsymbol{i}'}) = 1$ if $\boldsymbol{i}$ and $\boldsymbol{i}'$ are not equal in any components.

From Lemma 1, the coefficient of σ_u^2 in $\mathbb{E}_{\mathrm{RE}}(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*))$ is

$$\begin{aligned}
&\frac{1}{N^2} \sum_{\boldsymbol{i}} \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \Big(\mathbf{1}_{i_u = i'_u} - \frac{N_{\boldsymbol{i},u}}{N} - \frac{N_{\boldsymbol{i}',u}}{N} + \frac{\nu_u}{N} \Big) (1+\tau^2)^{|M_{\boldsymbol{i}\boldsymbol{i}'}|} \\
&= \frac{1}{N^2} \sum_{\boldsymbol{i}} \sum_{\boldsymbol{i}'} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \Big(\mathbf{1}_{i_u = i'_u} - \frac{2N_{\boldsymbol{i},u}}{N} + \frac{\nu_u}{N} \Big) (1+\tau^2)^{|M_{\boldsymbol{i}\boldsymbol{i}'}|} \\
&= \frac{1}{N^2} \sum_{k=0}^{r} (1+\tau^2)^k \sum_{\boldsymbol{i}} \sum_{\boldsymbol{i}'} \mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}| = k} Z_{\boldsymbol{i}} Z_{\boldsymbol{i}'} \Big(\mathbf{1}_{i_u = i'_u} - \frac{2N_{\boldsymbol{i},u}}{N} + \frac{\nu_u}{N} \Big) \\
&= \frac{1}{N} \sum_{k=0}^{r} (1+\tau^2)^k \big(\nu_{k,u} - 2\widetilde{\nu}_{k,u} + \rho_k \nu_u \big).
\end{aligned}$$

□

Next we establish an interpretable approximation to the Bayesian pigeonhole bootstrap variance, using the quantity $\epsilon = \max_{\boldsymbol{i}} \max_j N_{\boldsymbol{i},\{j\}}/N$ which is small unless the data are extremely imbalanced.

Theorem 6. *In the random effects model* (1)*, the gain coefficient* (14) *for* $u \neq \varnothing$ *in the product reweighted bootstrap is*

$$\gamma_u = \nu_u[(1+\tau^2)^{|u|} - 1 + \Theta_u \varepsilon] + \sum_{v \supsetneq u} (1+\tau^2)^{|v|} (\tau^2)^{|v-u|} \nu_v \tag{15}$$

where $|\Theta_u| \le (1+\tau^2)((1+\tau^2)^r - 1)/\tau^2$*. For* $\tau^2 = 1$*,*

$$\gamma_u = \nu_u[2^{|u|} - 1 + \Theta_u \varepsilon] + \sum_{v \supsetneq u} 2^{|v|} \nu_v,$$

where $|\Theta_u| \le 2^{r+1} - 2$*.*

Proof. The second claim follows immediately from the first which we now prove. We will approximate $\gamma_u = \sum_{k=0}^{r}(1+\tau^2)^k(\nu_{k,u} - 2\widetilde{\nu}_{k,u} + \rho_k\nu_u)$. First

$$\begin{aligned}
\sum_{k=0}^{r}(1+\tau^2)^k\nu_{k,u} &= \frac{1}{N}\sum_{k=0}^{r}(1+\tau^2)^k\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}\mathbf{1}_{i_u=i'_u}\\
&= \frac{1}{N}\sum_{w\supseteq u}(1+\tau^2)^{|w|}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbf{1}_{M_{ii'}=w}\\
&= \frac{1}{N}\sum_{w\supseteq u}(1+\tau^2)^{|w|}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\sum_{v\supseteq w}(-1)^{|v-w|}\mathbf{1}_{i_w=i'_w}\\
&= \sum_{w\supseteq u}(1+\tau^2)^{|w|}\sum_{v\supseteq w}(-1)^{|v-w|}\nu_v.
\end{aligned}$$

Writing $w\in[u,v]$ for $u\subseteq w\subseteq v$,

$$\begin{aligned}
&\sum_{w\supseteq u}(1+\tau^2)^{|w|}\sum_{v\subseteq w}(-1)^{|v-w|}\nu_v\\
&= \sum_{v\supseteq u}\nu_v\sum_{w\in[u,v]}(1+\tau^2)^{|w|}(-1)^{|v-w|}\\
&= \sum_{v\supseteq u}\nu_v\sum_{\ell=0}^{|v-u|}\binom{|v-u|}{\ell}(-1)^{\ell}(1+\tau^2)^{|v|-\ell}\\
&= \sum_{v\supseteq u}\nu_v(1+\tau^2)^{|v|}(\tau^2)^{|v-u|}.
\end{aligned}$$

For the other parts of γ_u, we use quantities θ that satisfy bounds $0\le\theta\le 1$. There are several such quantities, distinguished by subscripts, and defined at their first appearance. First, we have the bounds

$$\frac{N_{\boldsymbol{i},0}}{N} = 1 - r\theta_{\boldsymbol{i},0}\epsilon, \quad \text{and} \quad \frac{N_{\boldsymbol{i},k}}{N} = \theta_{\boldsymbol{i},k}\epsilon, \quad 1\le k\le r. \tag{25}$$

Next, for $u\neq\varnothing$,

$$\widetilde{\nu}_{0,u} = \frac{1}{N^2}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},u}N_{\boldsymbol{i},0} = \frac{1}{N}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},u}(1-r\theta_{\boldsymbol{i},0}\epsilon) = \nu_u(1-r\theta_{0,u}\epsilon)$$

and for $k=1,\dots,r$

$$\widetilde{\nu}_{k,u} = \frac{1}{N^2}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},u}N_{\boldsymbol{i},k} = \frac{1}{N}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},u}\theta_{\boldsymbol{i},k}\epsilon = \nu_u\theta_{k,u}\epsilon,$$

Turning to ρ_k,

$$\begin{aligned}
\rho_0 &= \frac{1}{N^2}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},0} = \frac{1}{N}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}(1-\epsilon r\theta_{\boldsymbol{i},0}) = 1-\epsilon r\theta_0, \quad \text{and}\\
\rho_k &= \frac{1}{N^2}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}N_{\boldsymbol{i},k} = \frac{1}{N}\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}\theta_{\boldsymbol{i},k}\epsilon = \theta_k\epsilon, \quad k=1,\dots,r.
\end{aligned}$$

Now $-2\widetilde{\nu}_{0,u} + \rho_0\nu_u = -\nu_u + \nu_u(2\theta_{0,u} - \theta_0)r\epsilon$ and

$$\sum_{k=1}^{r}(1+\tau^2)^k(-2\widetilde{\nu}_{k,u} + \rho_k\nu_u) = \nu_u\sum_{k=1}^{r}(1+\tau^2)^k(\theta_k - 2\theta_{k,u})\epsilon$$

Therefore

$$\gamma_u = \nu_u\big((1+\tau^2)^{|u|} - 1 + \Theta_u\epsilon\big) + \sum_{v\supsetneq u}\nu_v(1+\tau^2)(\tau^2)^{|v-u|}, \quad \text{where}$$

$$\Theta_u = \sum_{k=1}^{r}(1+\tau^2)^k(\theta_k - 2\theta_{k,u}).$$

The proof follows because $-1 \le \theta_k - 2\theta_{k,u} \le 1$ and $\sum_{k=1}^{r}(1+\tau^2)^k = (1+\tau^2)((1+\tau^2)^r - 1)/\tau^2$. □

Theorem 7. *For the random effects model* (1) *and the product reweighted bootstrap with* $\tau^2 = 1$*, the gain coefficient for nonempty* $u \subseteq [r]$ *satisfies*

$$2^{|u|} - 1 - (2^{r+1} - 2)\epsilon < \frac{\gamma_u}{\nu_u} \le 2^{|u|}(1+2\eta)^{|v-u|} - 1 + (2^{r+1} - 2)\epsilon.$$

If there exist m *and* M *with* $0 < m \le \sigma_u^2 \le M < \infty$ *for all* $u \ne \varnothing$*, then*

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = 1 + O(\eta + \epsilon).$$

Proof. From Theorem 6

$$\begin{aligned}\frac{\gamma_u}{\nu_u} &\le -1 + \sum_{v\supseteq u} 2^{|v|}\eta^{|v-u|} + (2^{r+1} - 2)\epsilon \\ &= 2^{|u|}(1+2\eta)^{|v-u|} - 1 + (2^{r+1} - 2)\epsilon,\end{aligned}$$

and then using $\nu_v > 0$,

$$\frac{\gamma_u}{\nu_u} > 2^{|u|} - 1 - (2^{r+1} - 2)\epsilon.$$

For the second claim, small η means that the variance is dominated by contributions $\sigma^2_{\{j\}}$ for which $\gamma_{\{j\}} \approx \nu_{\{j\}}$. Now

$$\sum_{|u|=1}\gamma_u\sigma_u^2 = \sum_{|u|=1}\nu_u\sigma_u^2[1 + O(\eta+\epsilon)]$$

where the constant in $O(\cdot)$ can depend on r, and

$$\sum_{|u|>1}\gamma_u\sigma_u^2 = \sum_{|u|>1}\nu_u\sigma_u^2[2^{|u|} + O(\eta+\epsilon)] = O(\eta)\sum_{|u|=1}\nu_u\sigma_u^2.$$

Similarly $\sum_{|u|>1}\gamma_u\sigma_u^2 = O(\eta)\sum_{|u|=1}\nu_u\sigma_u^2$. Therefore

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = \frac{(1+O(\eta+\epsilon))\sum_{|u|=1}\nu_u\sigma_u^2}{(1+O(\eta))\sum_{|u|=1}\nu_u\sigma_u^2} = 1+O(\eta+\epsilon).$$

□

Proof of Theorems 8 through 11.

Here we prove the theorems for the heteroscedastic case. We begin with a lemma.

Lemma 2. *Let $X_{\boldsymbol{i}}$ follow the heteroscedastic random effects model* (17) *and let $Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} - \bar{X}$. Then*

$$\mathbb{E}_{\mathrm{RE}}(X_{\boldsymbol{i}}X_{\boldsymbol{i}'}) = \mu^2 + \sum_{u\neq\varnothing}\sigma^2_{\boldsymbol{i},u}\mathbf{1}_{i_u=i'_u} \tag{26}$$

and

$$\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}) = \sum_{u\neq\varnothing}\Big(\mathbf{1}_{i_u=i'_u}\sigma^2_{\boldsymbol{i},u} - \nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u} - \nu_{\boldsymbol{i}',u}\sigma^2_{\boldsymbol{i}',u} + \overline{\nu_u\sigma_u^2}\Big). \tag{27}$$

Proof. Equation (26) follows directly just as the analogous expression did in Lemma 1. Once again, expanding $Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'}$ yields

$$X_{\boldsymbol{i}}X_{\boldsymbol{i}'} - \frac{1}{N}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}X_{\boldsymbol{i}}X_{\boldsymbol{i}''} - \frac{1}{N}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}X_{\boldsymbol{i}'}X_{\boldsymbol{i}''} + \frac{1}{N^2}\sum_{\boldsymbol{i}''}\sum_{\boldsymbol{i}'''}Z_{\boldsymbol{i}''}Z_{\boldsymbol{i}'''}X_{\boldsymbol{i}''}X_{\boldsymbol{i}'''}.$$

and we may assume that $\mu = 0$ while proving (27). Now

$$\mathbb{E}_{\mathrm{RE}}\Big(\frac{1}{N}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}'}X_{\boldsymbol{i}}X_{\boldsymbol{i}'}\Big) = \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}'}\mathbf{1}_{i_u=i'_u}\sigma^2_{\boldsymbol{i},u} = \sum_{u\neq\varnothing}\sum_{\boldsymbol{i}}\sigma^2_{\boldsymbol{i},u}\nu_{\boldsymbol{i},u},$$

and

$$\begin{aligned}\mathbb{E}_{\mathrm{RE}}\Big(\frac{1}{N^2}\sum_{\boldsymbol{i}''}\sum_{\boldsymbol{i}'''}Z_{\boldsymbol{i}''}Z_{\boldsymbol{i}'''}X_{\boldsymbol{i}''}X_{\boldsymbol{i}'''}\Big) &= \frac{1}{N^2}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}''}\sum_{\boldsymbol{i}'''}Z_{\boldsymbol{i}''}Z_{\boldsymbol{i}'''}\mathbf{1}_{i''_u=i'''_u}\sigma^2_{\boldsymbol{i}'',u}\\ &= \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}''}Z_{\boldsymbol{i}''}\sigma^2_{\boldsymbol{i}'',u}\nu_{\boldsymbol{i}'',u}\\ &= \sum_{u\neq\varnothing}\overline{\nu_u\sigma_u^2},\end{aligned}$$

which together establish (27). □

Theorem 8. *In the heteroscedastic random effect model* (17)

$$\mathrm{Var}(\bar{X}) = \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}}\nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u}. \tag{28}$$

Proof. The proof is very similar to that of Theorem 1. □

Theorem 9. *In the heteroscedastic random effects model* (17)

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u}, \tag{19}$$

where

$$\gamma_{\boldsymbol{i},u} = \sum_{k=0}^{r}(1+\tau^2)^k(\nu_{\boldsymbol{i},k,u} - 2\nu_{\boldsymbol{i},k}\nu_{\boldsymbol{i},u} + \overline{\nu_k}\nu_{\boldsymbol{i},u}). \tag{20}$$

Proof. We begin along the same lines as Theorem 3 and find that

$$\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big) = \frac{1}{N^2}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}\mathbb{E}_{\mathrm{RE}}(Y_{\boldsymbol{i}}Y_{\boldsymbol{i}'})\mathbb{E}_{\mathrm{PW}}(W_{\boldsymbol{i}}W_{\boldsymbol{i}'}).$$

As in Theorem 5, $\mathbb{E}_{\mathrm{PW}}(W_{\boldsymbol{i}}W_{\boldsymbol{i}'}) = (1+\tau^2)^{|M_{\boldsymbol{i}\boldsymbol{i}'}|}$.

From Lemma 2,

$$\begin{aligned}\mathbb{E}_{\mathrm{RE}}(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)) = \frac{1}{N^2}\sum_{u\neq\varnothing}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}}Z_{\boldsymbol{i}'}(1+\tau^2)^{|M_{\boldsymbol{i}\boldsymbol{i}'}|}\\ \times\Big(\mathbf{1}_{i_u=i'_u}\sigma^2_{\boldsymbol{i},u} - \nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u} - \nu_{\boldsymbol{i}',u}\sigma^2_{\boldsymbol{i}',u} + \overline{\nu_u\sigma^2_u}\Big).\end{aligned} \tag{29}$$

The contribution from the last term in the parentheses of (29) is

$$\frac{1}{N}\sum_{u\neq\varnothing}\overline{\nu_u\sigma^2_u}\sum_{k=0}^{r}(1+\tau^2)^k\sum_{\boldsymbol{i}}Z_{\boldsymbol{i}}\nu_{\boldsymbol{i},k} = \sum_{u\neq\varnothing}\overline{\nu_u\sigma^2_u}\sum_{k=0}^{r}(1+\tau^2)^k\overline{\nu_k}.$$

Therefore the coefficient of $\sigma^2_{\boldsymbol{i},u}$, in $\mathbb{E}_{\mathrm{RE}}(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*))$ (when $Z_{\boldsymbol{i}}=1$) is

$$\begin{aligned}&\frac{1}{N^2}\sum_{\boldsymbol{i}'}Z_{\boldsymbol{i}'}\sum_{k=0}^{r}\mathbf{1}_{|M_{\boldsymbol{i}\boldsymbol{i}'}|=k}(1+\tau^2)^k\big(\mathbf{1}_{i_u=i'_u} - 2\nu_{\boldsymbol{i},u}\big) + \frac{\nu_{\boldsymbol{i},u}}{N}\sum_{k=0}^{r}(1+\tau^2)^r\overline{\nu_k}\\ &= \frac{1}{N}\sum_{k=0}^{r}(1+\tau^2)^k(\nu_{\boldsymbol{i},k,u} - 2\nu_{\boldsymbol{i},k}\nu_{\boldsymbol{i},u} + \overline{\nu_k}\nu_{\boldsymbol{i},u}).\end{aligned}$$

□

Theorem 10. *In the heteroscedastic random effects model* (17)*, the gain coefficient* $\gamma_{\boldsymbol{i},u}$ *of* (20) *for* $Z_{\boldsymbol{i}}=1$ *and* $u\neq\varnothing$*, in the product reweighted bootstrap is*

$$\gamma_{\boldsymbol{i},u} = \nu_{\boldsymbol{i},u}[(1+\tau^2)^{|u|} - 1 + \Theta_u\varepsilon] + \sum_{v\supsetneq u}(1+\tau^2)^{|v|}(\tau^2)^{|v-u|}\nu_{\boldsymbol{i},v}$$

where $|\Theta_u| \le (1+\tau^2)((1+\tau^2)^r - 1)/\tau^2$*. For* $\tau^2 = 1$

$$\gamma_{\boldsymbol{i},u} = \nu_{\boldsymbol{i},u}[2^{|u|} - 1 + \Theta_u\varepsilon] + \sum_{v\supsetneq u}2^{|v|}\nu_{\boldsymbol{i},v}$$

where $|\Theta_u| \le 2^{r+1} - 2$*.*

Proof. From Theorem 9, $\gamma_{\boldsymbol{i},u} = \sum_{k=0}^{r}(1+\tau^2)^k(\nu_{\boldsymbol{i},k,u} - 2\nu_{\boldsymbol{i},k}\nu_{\boldsymbol{i},u} + \overline{\nu_k}\nu_{\boldsymbol{i},u})$. The proof is similar to that of Theorem 6, so we summarize the steps. First

$$\sum_{k=0}^{r}(1+\tau^2)^k\nu_{\boldsymbol{i},k,u} = \sum_{v\supseteq u}\nu_{\boldsymbol{i},v}(1+\tau^2)^{|v|}(\tau^2)^{|v-u|}.$$

Next, $\nu_{\boldsymbol{i},0} = 1 - r\theta_{\boldsymbol{i},0}\epsilon$ and $\overline{\nu_0} = 1 - r\theta_0$, while for $k \geq 1$, $\nu_{\boldsymbol{i},k} = \theta_{\boldsymbol{i},k}\epsilon$ and $\overline{\nu_k} = \theta_k\epsilon$. Here, all of the θ's are in the interval $[0,1]$. The result follows as in Theorem 6. □

Theorem 11. *For the heteroscedastic random effects model* (17)*, assume that there exist m and M with $0 < m \leq \sigma^2_{\boldsymbol{i},u} \leq M < \infty$. Then the product reweighted bootstrap with $\tau^2 = 1$, satisfies*

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = 1 + O(\eta+\epsilon).$$

Proof. First we show that main effects dominate. For $|u| > 1$,

$$\begin{aligned}
\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u} &\leq M\sum_{\boldsymbol{i}}\nu_{\boldsymbol{i},u}(2^{|u|} - 1 + 2^{r+1}\epsilon) + \sum_{v\supsetneq u}2^{|v|}\nu_{\boldsymbol{i},v}\\
&= M\left(\nu_u(2^{|u|} - 1 + 2^{r+1}\epsilon) + \sum_{v\supsetneq u}2^{|v|}\nu_v\right)\\
&= (2^{|u|}-1)M\nu_u(1+O(\epsilon+\eta))\\
&= O(\eta)\max_{1\leq j\leq r}\nu_{\{j\}},
\end{aligned}$$

and similarly $\sum_{\boldsymbol{i}}\nu_{\boldsymbol{i},u}\sigma^2_{\boldsymbol{i},u} = O(\eta)\max_{1\leq j\leq r}\nu_{\{j\}}$. For $u = \{j\}$,

$$\begin{aligned}
\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i},\{j\}}\sigma^2_{\boldsymbol{i},\{j\}} &\geq m\sum_{\boldsymbol{i}}\nu_{\boldsymbol{i},\{j\}}(1-2^{r+1}\epsilon)\\
&= m\nu_{\{j\}}(1+O(\epsilon)).
\end{aligned}$$

Therefore

$$\frac{\mathbb{E}_{\mathrm{RE}}\big(\widetilde{\mathrm{Var}}_{\mathrm{PW}}(\bar{X}^*)\big)}{\mathrm{Var}(\bar{X})} = \frac{\sum_{\boldsymbol{i}}\sum_{j=1}^{r}\gamma_{\boldsymbol{i},\{j\}}\sigma^2_{\boldsymbol{i},\{j\}}}{\sum_{\boldsymbol{i}}\sum_{j=1}^{r}\nu_{\boldsymbol{i},\{j\}}\sigma^2_{\boldsymbol{i},\{j\}}}(1+O(\eta+\epsilon)).$$

Next we show that the main effects are properly estimated

$$\begin{aligned}
\sum_{\boldsymbol{i}}\sum_{j=1}^{r}|\gamma_{\boldsymbol{i},\{j\}} - \nu_{\boldsymbol{i},\{j\}}|\sigma^2_{\boldsymbol{i},\{j\}} &\leq M\sum_{\boldsymbol{i}}\sum_{j=1}^{r}|\gamma_{\boldsymbol{i},\{j\}} - \nu_{\boldsymbol{i},\{j\}}|\\
&\leq M\sum_{\boldsymbol{i}}\sum_{j=1}^{r}\nu_{\boldsymbol{i},\{j\}}(2^{r+1}\epsilon + 3^r\eta)\\
&= \sum_{j=1}^{r}\nu_{\{j\}}O(\eta+\epsilon),
\end{aligned}$$

while $\sum_{\boldsymbol{i}} \sum_{j=1}^{r} \nu_{\boldsymbol{i},\{j\}} \sigma^2_{\boldsymbol{i},\{j\}} \geq m \sum_{j=1}^{r} \nu_{\{j\}}$. $\square$