Turning Long and Short Return Histories into Equal Histories: A Better Way to Backfill Returns

Yindeng Jiang and Doug Martin

August, 2016

Abstract

Standard quantitative portfolio analysis techniques, including mean-variance analysis, historical risk and performance estimation, and various portfolio optimization techniques, implicitly require all assets under consideration having the same length of return histories. Unfortunately, it is often the case that certain assets have shorter return histories than others. We present a method, which is a significant improvement over an existing technique, for efficiently backfilling missing returns while accounting for non-normal distributions.

Keywords: short/unequal return histories, backfilling returns, non-normality

1 Introduction

Standard quantitative portfolio analysis techniques, including mean-variance analysis, historical risk and performance estimation, and various portfolio optimization techniques, implicitly require all assets under consideration having the same length of return histories. Unfortunately, it is often the case that certain assets have shorter return histories than others. For example, different indexes often have different inception dates, and a group of investment funds are unlikely to start their operation at the same time.

An *ad hoc* procedure used by analysts to address this challenge is to simply discard some of the early return histories so that the remaining return histories all share the same length, in other words, truncating all return histories to the shortest common history. Stambaugh

[1997] argues that valuable information is lost when return histories are truncated, especially considering "not only do those discarded returns provide additional information about the longer-history assets, but they generally provide information about the shorter-history assets as well." Hence Stambaugh [1997] proposes a technique that can effectively utilize all available return histories, which is essentially the maximum likelihood estimation (MLE) under multivariate normality based on the work of Anderson [1957]. However, financial time series data are notoriously non-normal, consequently a few methods have appeared in the literature that account for non-normality, most notably the class of factor-model Monte Carlo (FMMC) methods developed by Jiang [2009] and Jiang and Martin [2015], and the backfilling method of Page [2013], which can find its origins in the multiple imputation of Rubin [1976].

As we will explain later, while Page's method is intuitive, its averaging process creates both theoretical and practical challenges, especially in complex applications such as portfolio optimization. We present a special version of FMMC (mentioned in Jiang [2009, Sec. 6.2]) that we will refer to as *combined backfilling*, which is also closely related to Page's method. Basically, by combining the backfilled samples generated by Page's method, we create a method that can account for non-normality in a way similar to Page's method, and in the mean time inherits the implementation simplicity and computational efficiency of FMMC. Interestingly, the relationship between combined backfilling and Page's method is analogous to that between the fractionally weighted imputation of Fay [1996] and the multiple imputation of Rubin [1976].

In this paper we focus on the intuition and application of combined backfilling. Some theoretical arguments are provided in Jiang [2009]. The rest of the paper is organized as follows. Section 2 presents a minimal numerical example illustrating the mechanism of Page's method and combined backfilling. Section 3 discusses the challenges with Page's method and the advantages of combined backfilling. In Section 4, we replicate the numerical examples in Page [2013] to assess the performance of combined backfilling in capturing non-normality in the data. Section 5 includes a brief description of the iterative procedure for applying combined backfilling to more than two assets. Section 6 concludes the paper.

2 A minimal numerical example

We provide a minimal numerical example illustrating Page's method and contrast it with combined backfilling. Consider two assets A and B with only 4 months of returns and B only started since Month 2:

	A	B
1	0.05	
2	0.10	0.31
3	0.09	0.14
4	0.18	0.29

Page's method begins by regressing B on A using the common 3 months of returns, via ordinary least squares (OLS). This results in the following coefficients and residuals:

$$\hat{\alpha} = 0.13,\ \hat{\beta} = 0.93,\ \hat{\boldsymbol{\epsilon}} = (0.09, -0.08, -0.01) \tag{1}$$

Next we calculate the predicted value of B's Month 1 return based on the regression, ie, $\hat{\alpha} + \hat{\beta} \cdot 0.05 = 0.18$. To take into account of the uncertainty with the missing values, we draw a residual (with replacement) from the 3 residuals above and add it to the predicted value, which together with the original sample constitutes a backfilled sample. This last step is generally repeated many times (say, 10,000 times) to obtain a distribution of the backfilled samples. In the current example, it is clear that a backfilled sample will have to be one of the following three:

	A	B
1	0.05	0.18 + 0.09
2	0.10	0.31
3	0.09	0.14
4	0.18	0.29

	A	B
1	0.05	0.18 + (-0.08)
2	0.10	0.31
3	0.09	0.14
4	0.18	0.29

	A	B
1	0.05	0.18 + (-0.01)
2	0.10	0.31
3	0.09	0.14
4	0.18	0.29

So actually simply considering these 3 backfilled samples is enough to get the full distribution of the backfilled samples. However, as we explain later, this kind of exhaustive simulation is generally not practical.

Any estimation of interest (such as estimating volatility) is then done on each backfilled sample, and the average across backfilled samples is used as the final estimate. In essence, Page's method accounts for the uncertainty with the missing values by simulating many possible realizations of the missing values.

While this is intuitive, the averaging process of Page's method creates both theoretical and practical challenges as we discuss in the next section. Combined backfilling eliminates the averaging step by simply combining the backfilled samples into a large sample, and then any estimation of interest is done only once on this single large sample. In the current example, combined backfilling results in the combined backfilled sample as shown in Table 1.

Table 1: Combined backfilled sample

	A	B
1	0.05	0.18 + 0.09
2	0.10	0.31
3	0.09	0.14
4	0.18	0.29
5	0.05	0.18 + (-0.08)
6	0.10	0.31
7	0.09	0.14
8	0.18	0.29
9	0.05	0.18 + (-0.01)
10	0.10	0.31
11	0.09	0.14
12	0.18	0.29

3 Challenges with Page's method

In this section we discuss the theoretical and practical challenges in applying Page's method due to its averaging process. The same challenges do not apply to combined backfilling since no averaging of estimates is needed.

First of all, estimation with Page's method can be ambiguous. Consider, for example, one wishes to estimate vol with Page's method.[1] Clearly one can calculate the vol from

[1]This is a hypothetical example since vol can be estimated with Stambaugh's method, so no residual

each backfilled sample and then take the average. However, a sensible alternative would be to calculate the variance from each backfilled sample and then take the square root of the average variance across backfilled samples. The two approaches would in general give two different results, and the difference does not disappear even when one takes the number of backfilled samples to infinity, or equivalently, uses the exhaustive simulation we discussed above. (Statistically, this is referred to as not being parameter-transformation invariant.) For instance, using the minimal numerical example above, the two approaches give the following two estimates of vol: 0.0768 and 0.0775, respectively.[2] While the difference between the two estimates is usually small in practice, the mere existence of the difference reflects a theoretical challenge with the averaging process of Page's method. In fact, while Page's method is seemingly an extension to Stambaugh's method that accounts for higher moments, its estimates of the first two moments (ie, means and covariances) do not necessarily agree with Stambaugh's estimates, since the latter are MLEs under normality that are parameter-transformation invariant. On the other hand, as shown in Appendix A, combined backfilling is always consistent with Stambaugh's method for estimating the first two moments, thus it is a true extension to Stambaugh's method.

A more challenging situation arises in portfolio optimization. Since the advantage of Page's method over Stambaugh's method is its ability to account for higher moments, it is appropriate to consider portfolio optimization techniques that also account for higher moments, such as the mean-CVaR optimization of Rockafellar and Uryasev [2000] or the full-scale optimization of Adler and Kritzman [2007]. In either case we need a full sample to obtain an efficient frontier, which we can do with each backfilled sample. However, how should we then average the resulting efficient frontiers? Since each point on an efficient frontier represents an optimal portfolio with respect to a given level of risk aversion, a sensible approach would be to average the optimal portfolios with the same level of risk aversion. For example, for a completely risk averse investor, he would choose the average minimum-risk portfolios across the backfilled samples. The question is, does the average portfolio indeed represent an optimal portfolio? Not necessarily. Consider, for example,

resampling with either Page's method or combined backfilling is necessary. See next section for a more practical example.

[2]Here the vol and the variance are estimated as the population standard deviation and the population variance, respectively, to be consistent with MLEs, but the sample standard deviation and the sample variance show similar discrepancies.

the investor imposes a constraint that no more than 10 assets should be included in his portfolios, then averaging multiple portfolios that satisfy that constraint will likely result in a portfolio that has more than 10 assets, hence it is not even feasible (let alone optimal) from the investor's perspective.[3] Therefore, it becomes clear that the root problem with Page's method is that the estimates from backfilled samples may not be "average-able," so perhaps the remedy is to transform the estimates into something that is average-able before taking the averages, but that creates a challenge of its own and it will be difficult for practitioners to implement. On the other hand, with combined backfilling it is straightforward to apply any portfolio optimization techniques that require a full sample since no "averaging" is needed.

Another advantage of combined backfilling is its computational efficiency. Note a backfilled sample is obtained by drawing k residuals (with replacement) from the n residuals from the regression (k is the number of missing values, and n is the length of the common histories), by the multiplication principle, the number of possible backfilled samples is n^k. Therefore, the exhaustive simulation for Page's method would require generating all n^k possible backfilled samples and performing the estimation of interest n^k times. This is not practical since n^k is usually prohibitively large. For example, if the short history is 3 years of monthly returns and the long history is 10 years of monthly returns, we have

$$n^k = 36^{120-36} = 5.363028 \times 10^{130} \tag{2}$$

Even when n^k is of a more reasonable size, the estimation can be computationally expensive (such as the portfolio optimization techniques above), thus repeating the estimation n^k times can still be a computationally challenging task. On the other hand, when we combine all n^k possible backfilled samples into a single sample, the resulting seemingly large sample is actually consisted of n^{k-1} replicates of a much smaller sample of size $n(n+k)$. This can be seen by noting, in the large sample, each observed value is repeated n^k times, and each unique backfilled value is repeated n^{k-1} times. Hence the smaller sample, which we will refer to as the *fully combined backfilled sample* (FCBS), can be viewed as the combination of n special backfilled-samples, each of which is backfilled with a single residual that is applied to all k missing values. Therefore, with combined backfilling, we only need to backfill n

[3]Basically, since taking the average is a linear operation, any non-linear constraint on the portfolio weights may be violated by taking the average of portfolios that do satisfy the constraint.

times using the n residuals deterministically, combine these backfilled samples to form the FCBS, and then perform any estimation of interest once on the FCBS. Since the procedure is exhaustive, no simulation error is incurred. Moreover, the size of the FCBS, $n(n+k)$, is generally manageable computationally. Using the same example as above,

$$n(n+k) = 36 \cdot 120 = 4320 \tag{3}$$

4 A two-asset example

Does combined backfilling work, ie, can it properly account for non-normality? Below we replicate the two-asset examples in Page [2013]. The results suggest that combined backfilling performs better or similar to Page's method in capturing non-normality in the data.

The data sample comprises monthly returns on the Wilshire 5000 Total Market Index (W5000) from January 1971 to May 2011 and on the MSCI Emerging Markets Index (MSCI EM) from January 1988 to May 2011. Regressing MSCI EM on W5000 via OLS using the common return histories starting from January 1988 produces:

$$\hat{\alpha} = 0.0035, \ \hat{\beta} = 1.1, \tag{4}$$

and 281 residuals. Figure 1 plots the histogram of the residuals superimposed with a normal density curve. It appears that the residuals might have a fat left tail, but it is hardly conclusive. Figure 2 shows the normal quantile-quantile (QQ) plot of the residuals, which makes it abundantly clear that the residuals have fat tails on both the left and the right side of the distribution. This suggests resampling from the residuals is important in capturing the non-normality in the data, providing support for both Page's method and combined backfilling.

As in Page [2013], we consider the estimation of skewness and kurtosis based on backfilling the missing returns of MSCI EM. For Page's method, 10,000 backfilled samples are randomly generated. We calculate the skewness and kurtosis of MSCI EM using the backfilled returns from January 1971 to December 1987 in each backfilled sample,[4] and then take the average

[4]To be consistent with Page [2013] and to directly assess the features of the backfilled returns, we estimate the skewness and kurtosis using only the backfilled return as opposed to the full sample for MSCI EM. In practical applications the latter will usually be preferred.

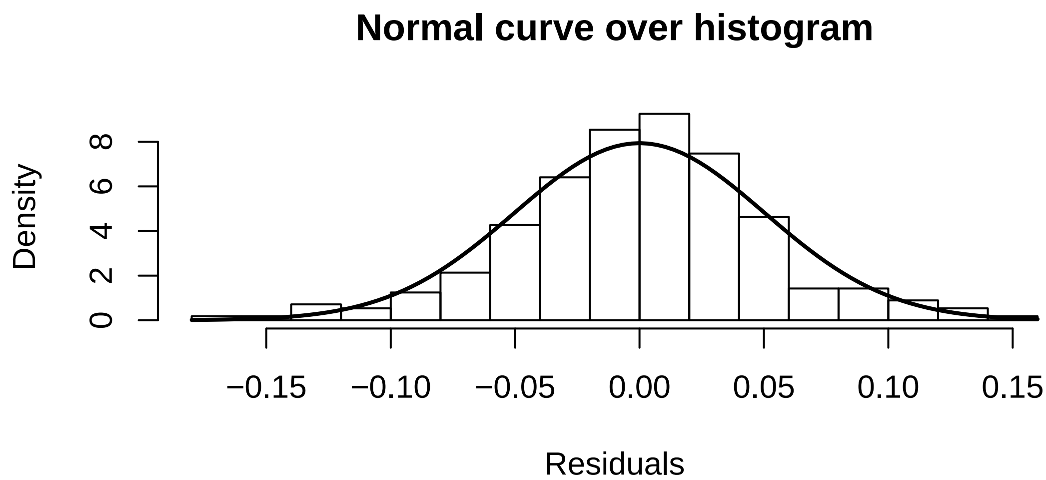

Figure 1: Distribution of residuals

skewness and kurtosis across 10,000 samples. For combined backfilling, as explained above, we only need to combine the 281 backfilled returns paths for MSCI EM from January 1971 to December 1987, each backfilled with a single residual that is applied to all months during the 17 year period. The FCBS essentially comprises all possible realizations of the MSCI EM return on a random month during the period. The skewness and kurtosis of this FCBS, along with the estimates from Page's method, are shown in Table 2. It shows that both Page's method and combined backfilling are able to capture the negative skewness and higher than normal kurtosis in MSCI EM returns, and the differences between the two methods are likely too small to be statistically significant.

Table 2: Higher moments for MSCI EM, January 1971-December 1987

	Skewness	Kurtosis
Page's method	-0.20	3.82
Combined backfilling	-0.20	3.90
Normal distribution	0.00	3.00

Since

$$\text{Kurtosis} = \frac{\text{The 4th Sample Moment}}{\text{Sample Standard Deviation}^4}, \tag{5}$$

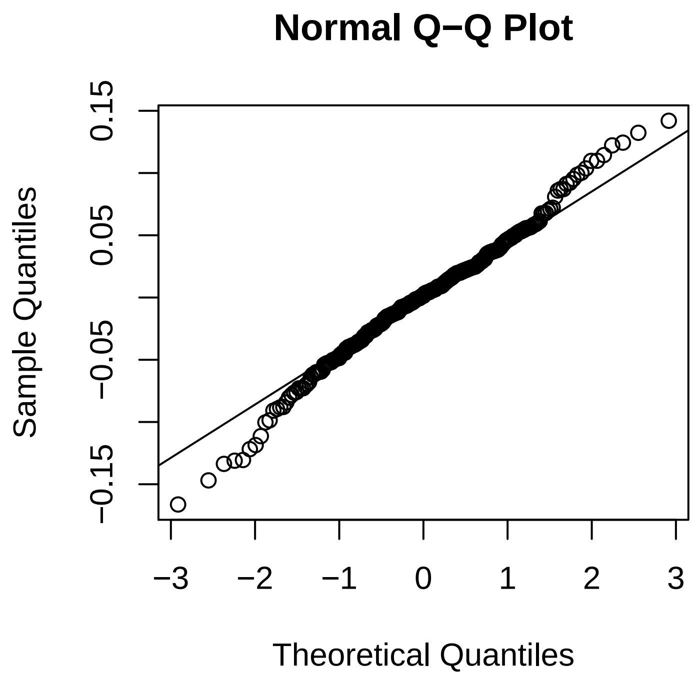

Figure 2: QQ plot of residuals

an alternative estimate of kurtosis with Page's method is obtained by first taking the average of the 4th sample moment and the sample standard deviation across the 10,000 backfilled samples, respectively, and then calculate the kurtosis as in (5). This yields an estimate of 3.88, which is meaningfully different than the estimate with Page's method in Table 2. This again illustrates the ambiguity when estimating with Page's method.

Next we replicate the experiment in Page [2013] to test how combined backfilling and Page's method perform in recovering the features of missing return histories. To do this, we remove the first 10 years of data for MSCI EM and use each method to backfill these missing data. Table 3 shows combined backfilling performs slightly better than Page's method, in the sense that it generates higher moments that are closer to the actual data.[5]

[5]The small differences between our results here and those in Table 3 of Page [2013] (especially for "Actual data") are likely due to: 1) different ways of calculating kurtosis (see Joanes and Gill [1998]; here we use the default implementation in the statistical software R), and 2) based on the caption of Table 3 of Page [2013], 11 rather than 10 years of data might have been removed from the MSCI EM histories.

Table 3: Higher moments for MSCI EM, January 1988-December 1997

	Skewness	Kurtosis
Page's method	-0.19	3.10
Combined backfilling	-0.20	3.21
Actual data	-0.26	3.76
Normal distribution	0.00	3.00

Finally we perform the simulation study as in Page [2013] to check if the above results are specific to the data under consideration. We generate 100,000 observations from a bivariate Student's t-distribution with a kurtosis of 6 (corresponding to degrees of freedom of 6 as well) for both assets and a correlation of 25%. We then remove 10,000 observations for one of the assets and backfill 5,000 times with each method. As before, we estimate the skewness and kurtosis using the backfilled missing data. Table 4 shows that both combined backfilling and Page's method are able to recover a significant portion of the excess kurtosis present in the removed data, and the two methods provide almost identical results in this case.[6] The flip in sign for the skewness estimates is likely due to the fact that the skewness for the actual data is close to 0 (note the t-distribution we are using has a skewness of 0).

Table 4: Higher moments for backfilled simulated data from fat-tailed t-distribution

	Skewness	Kurtosis
Page's method	0.03	5.21
Combined backfilling	0.03	5.21
Actual simulated data	-0.21	6.98
Normal distribution	0.00	3.00

5 Combined backfilling for multiple assets

We briefly describe the iterative procedure for applying combined backfilling when the number of assets is more than two, or more importantly, the number of the distinct lengths of the return histories is more than two. In such cases, we first order the assets from the longest

[6]There is a significant difference between our results here and those in Table 4 of Page [2013]. This is mostly likely due to simulation error since even with the large sample size we are using, we have experienced significant variations across simulation runs. Fortunately, the results are all directionally consistent.

histories to the shortest histories, and then group them based on the length of histories. Suppose there are m groups, and the ith group comprises k_i assets all with history length n_i ($n_1 > n_2 > \ldots > n_m$). Now we backfill the missing returns of the second group assets based on the returns of the first group assets similar to the two-asset case above (the simple OLS regression is replaced by multivariate OLS regression). Note we only need to backfill n_2 times, each with a single residual (vector) from the regression. We hold off on combining the backfilled samples until the end.

For each of these backfilled samples, we treat it as if it were the observed histories for the first two groups of assets. We then repeat the process to backfill the missing returns of the third group assets based on this backfilled sample of the first two groups of assets (note they now have the same length of histories). Since n_3 backfilled samples are generated for each of the n_2 backfilled samples above, now we have a total of $n_2 n_3$ backfilled samples for the first three groups of assets. The same process is then iteratively applied to group $4, \ldots,$ until group m, and we end up with a total of $n_2 n_3 \cdots n_m$ backfilled samples, each with length n_1. Combining all these backfilled samples yields a sample of size $n_1 n_2 \cdots n_m$, which is the FCBS that can be used for any estimation of interest.

6 Concluding remarks

We have presented an effective and easy to implement method for addressing the challenge of unequal return histories in portfolio analysis. By replicating the numerical examples in Page [2013], we have shown that combined backfilling performs better or similar to Page's method in capturing non-normality in the data. The examples focused on simple parameter estimation for non-normal distributions for ease of presentation, but the most advantage of combined backfilling is in complex applications such as portfolio optimization. As we have argued, Page's method simply does not yield valid results in certain complex applications. In other situations, while Page's method does provide useful results, they are probably best viewed as a heuristic approximation to combined backfilling.

A similar set of caveats to those in Page [2013] apply to combined backfilling:

- Combined backfilling does not magically create information by resampling from the residuals. Instead, it turns unequal histories into equal "histories" with minimum loss of information.

- The model assumes that data from the long sample contain valuable information. If data from the long sample are stale or contain misleading information, using the model will lead to deterioration of the analysis compared with simply using the short sample.
- The model assumes that betas between the existing factors and the missing factors do not change, which is not necessarily a realistic assumption. (To the extent betas change within the factors with long histories, however, the model should capture the effect. For example, if the correlations of rates and equities change when going back to the 1970s, then so long as we have data for rates and equities in the long sample, the model will simply propagate this change to the other related factors.)
- There is a limit to how much data can be backfilled. If a large portion of the data is missing, it might be better to simply use the shorter, common sample. Simulation analysis may help determine when information in the long sample is insufficient to warrant backfilling. In general, the higher the correlations and the more stable the distribution's moments, the more data that can be backfilled.
- We have focused on the statistical aspects of the methods and have not considered any economic constraints on the time series. For example, for such series as interest rates, negative values might be viewed as economically impossible (recent events in Europe might change this view though!), but there is nothing *a priori* in the algorithm to prevent backfilled rates from being negative. In short, there is no simple solution for all such problems, and analysts using these techniques should adapt them carefully to the problem at hand.

References

Timothy Adler and Mark Kritzman. Mean–variance versus full-scale optimisation: in and out of sample. *Journal of Asset Management*, 7(5):302–311, 2007.

T. W. Anderson. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, 52(278): 200–203, 1957.

R.E. Fay. Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, pages 490–498, 1996.

Y. Jiang. *Factor Model Monte Carlo Methods for General Fund-of-Funds Portfolio Management*. PhD thesis, University of Washington, Department of Statistics, 2009.

Yindeng Jiang and Doug Martin. Better risk and performance estimates with factor-model monte carlo. *Journal of Risk*, 17(5):1–38, 2015.

DN Joanes and CA Gill. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):183–189, 1998.

Sébastien Page. How to combine long and short return histories efficiently. *Financial Analysts Journal*, 69(1):45–52, 2013.

R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

D.B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Robert F. Stambaugh. Analyzing investments whose histories differ in length. *Journal of Financial Economics*, 45:285–331, 1997.

A Estimating means and covariances with combined backfilling

Consider two assets (x, y) with the following return histories:

$$x_1, \ldots, x_n, x_{n+1}, \ldots, x_N \tag{6}$$

$$y_1, \ldots, y_n \tag{7}$$

so $y_{n+1}, \ldots, y_N$ are missing. Let $\hat{\alpha}, \hat{\beta}$ be the OLS regression coefficients and $(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)$ be the residuals. We have

$$\sum_{i=1}^{n} \hat{\epsilon}_i = 0 \tag{8}$$

The normal distribution MLEs are given by (see Stambaugh [1997] or Anderson [1957])

$$
\begin{aligned}
\hat{\mu}_x &= \frac{1}{N}\sum_{i=1}^{N} x_i && (9)\\
\hat{\sigma}_x^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu}_x)^2 && (10)\\
\hat{\mu}_y &= \hat{\alpha} + \hat{\beta}\hat{\mu}_x && (11)\\
\hat{\sigma}_y^2 &= \hat{\beta}^2\hat{\sigma}_x^2 + \hat{\sigma}_\epsilon^2 && (12)\\
\hat{\sigma}_{x,y}^2 &= \hat{\beta}\hat{\sigma}_x^2 && (13)
\end{aligned}
$$

where

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2 \tag{14}$$

Recall the FCBS is given by

$$\{(x_{i,j}, y_{i,j}) : \; i = 1, \ldots, N; \; j = 1, \ldots n.\} \tag{15}$$

where $x_{i,j} = x_i, \; i = 1, \ldots, N; \; j = 1, \ldots n$, and $y_{i,j} = y_i, \; i = 1, \ldots, n; \; j = 1, \ldots n$, but for $i = n+1, \ldots, N$,

$$y_{i,j} = \hat{y}_i + \hat{\epsilon}_j, \; j = 1, \ldots, n. \tag{16}$$

where $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i, \; i = 1, \ldots, N$ are the fitted or predicted values of y.

The mean estimate of y based on the FCBS is then given by (note the sum of the residuals is 0)

$$
\begin{aligned}
\hat{\mu}_{y,\mathrm{FCBS}} &= \frac{1}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n} y_{i,j} && (17)\\
&= \frac{1}{nN}\left(\sum_{i=1}^{n} n y_i + \sum_{i=n+1}^{N} n\hat{y}_i\right) && (18)\\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i && (19)\\
&= \hat{\mu}_y && (20)
\end{aligned}
$$

and the variance estimate of y based on the FCBS is

$$\begin{aligned}
\hat{\sigma}^2_{y,\mathrm{FCBS}} &= \frac{1}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n}\left(y_{i,j}-\hat{\mu}_{y,\mathrm{FCBS}}\right)^2 && (21)\\
&= \frac{1}{nN}\left(\sum_{i=1}^{n} n\left(y_i-\hat{\mu}_y\right)^2+\sum_{i=n+1}^{N}\sum_{j=1}^{n}\left(\hat{y}_i+\hat{\epsilon}_j-\hat{\mu}_y\right)^2\right) && (22)\\
&= \frac{1}{nN}\left(\sum_{i=1}^{n} n\left(\hat{\alpha}+\hat{\beta}x_i+\hat{\epsilon}_i-\hat{\mu}_y\right)^2+\sum_{i=n+1}^{N}\sum_{j=1}^{n}\left(\hat{\alpha}+\hat{\beta}x_i+\hat{\epsilon}_j-\hat{\mu}_y\right)^2\right) && (23)\\
&= \frac{1}{nN}\left(\sum_{i=1}^{n} n\left(\hat{\beta}(x_i-\hat{\mu}_x)+\hat{\epsilon}_i\right)^2+\sum_{i=n+1}^{N}\sum_{j=1}^{n}\left(\hat{\beta}(x_i-\hat{\mu}_x)+\hat{\epsilon}_j\right)^2\right) && (24)
\end{aligned}$$

Note the residuals sum to 0 and they have zero correlation with $(x_1,\ldots,x_n)$, we have

$$\sum_{i=1}^{n}\hat{\beta}(x_i-\hat{\mu}_x)\hat{\epsilon}_i=0 \qquad (25)$$

and

$$\sum_{j=1}^{n}\hat{\beta}(x_i-\hat{\mu}_x)\hat{\epsilon}_j=0,\ \forall\, i \qquad (26)$$

so (24) evaluates to

$$\begin{aligned}
\hat{\sigma}^2_{y,\mathrm{FCBS}} &= \frac{1}{nN}\left(\sum_{i=1}^{n} n\left(\hat{\beta}^2(x_i-\hat{\mu}_x)^2+\hat{\epsilon}_i^2\right)+\sum_{i=n+1}^{N}\sum_{j=1}^{n}\left(\hat{\beta}^2(x_i-\hat{\mu}_x)^2+\hat{\epsilon}_j^2\right)\right) && (27)\\
&= \frac{1}{nN}\left(n\hat{\beta}^2\sum_{i=1}^{N}(x_i-\hat{\mu}_x)^2+n\cdot n\cdot\hat{\sigma}^2_\epsilon+(N-n)\cdot n\cdot\hat{\sigma}^2_\epsilon\right) && (28)\\
&= \hat{\beta}^2\hat{\sigma}^2_x+\hat{\sigma}^2_\epsilon && (29)\\
&= \hat{\sigma}^2_y && (30)
\end{aligned}$$

The covariance estimate between x and y based on the FCBS can be shown similarly to be exactly equal to $\hat{\sigma}^2_{x,y}$. Therefore, we have shown that combined backfilling provides the same estimates of means and covariances as those given by the normal distribution MLE or Stambaugh's method.