Consider $W_{ij} = Z_i^{\mathrm{T}}\beta + X_i + U_{ij}$. where $i = 1, \ldots, n$ and $j = 1, \ldots, m_j$. Assume:

- $Z_i \sim N(\mu_z, \sigma_z^2\mathbb{I})$
- $U_{ij} \sim N\{0, \sigma_U^2(Z_i)\}$ where $\sigma^2(\cdot)$ is an unknown smooth function
- $Cor(U_{ij}, U_{ik}) = 0$ for all $j \neq k$ (I think this assumption is reasonable)
- $X_i \sim N(0, \sigma_x^2)$

We have

$$Var(W_{ij}) = \beta^{\mathrm{T}}\sigma_z\beta + \sigma_x^2 + \sigma_u^2(Z_i) + \beta^{\mathrm{T}}\Sigma_{ZU}\beta + \beta^{\mathrm{T}}\Sigma_{ZX}\beta. \tag{1}$$

If we can estimate all the other components, we can solve for $\widehat{\sigma}_x^2$. If X and Z are uncorrelated, we have one less term to worry about since $\Sigma_{ZX} = 0$. This seems like a reasonable assumption to me. Is it possible to assume $\Sigma_{ZU} = 0$? I don't think Z and U can be independent if the variance of U is a function of Z.

Consider the replication differences, $W_{ij} - W_{ik} = U_{ij} - U_{ik}$, and note $Var(W_{ij} - W_{ik}) = Var(U_{ij}) + Var(U_{ik}) - 2Cor(U_{ij}, U_{ij}) = Var(U_{ij}) + Var(U_{ik}) = 2\sigma^2(Z_i)$. Then $1/\sqrt{2}(W_{ij} - W_{ik}) \sim N\{0, \sigma(Z_i)\}$ Now consider all permutations of the replication differences, the set $\{W_{ij} - W_{ik} | k \neq j\}$. This gives us $m_j(m_j - 1)$ observations for each Z_i to model the variance function, $\sigma_U^2(Z_i)$. Denote these differences as $D_{i\ell}$, for $\ell = 1, \ldots, m_j(m_j - 1)$ and $i = 1, \ldots, n$.

Ruppert et al. (1998) wrote a paper about variance estimation, and I'm stealing their general idea. I doubt the asymptotics hold. The paper is technical and I haven't gone through it in detail. I think we can use his idea though. You and Marie Davidian wrote a similar, though less general, paper: Davidian and Carroll (1987)

Make a new dataset. The responses variables is $D_{i\ell}$ and the explanatory variables are Z_i. This means we have to repeat each Z_i so there is a Z_i for each $D_{i\ell}$. So our new data is $\{(D_{11}, Z_1), (D_{12}, Z_1), \ldots D_{1m_1(m_1-1)}, Z_1), (D_{21}, Z_1), \ldots, (D_{nm_n(m_n-1)}, Z_n)\}$. This notation is a mess, sorry.

Regress $D_{i\ell} = m(Z_i) + \epsilon_i$ where $m(\cdot)$ is a specified or unspecified mean function which should hopefully be estimated very close to 0 and ϵ_i are iid $N\{0, \sigma^2(Z_i)\}$. Take the residuals

r_i and regression them against Z_i to get $\widehat{\sigma}^2(\cdot)$. That is $r_i = \sigma^2(Z_i) + \epsilon_i$. I think this will work with $Var(U_{ij}) = \sigma^2(X_i, Z_i)$ as well. We can fit $\widehat{\sigma}^2(\cdot)$ with multivariate spline model.

The good: I am actually able to get a decent estimate of σ_x^2. I am able to get reasonable estimates of $\widehat{\sigma}^2(\cdot)$, though there are problems at the tails of the data. We can estimate everything else in (1) with sample variance or covariance. All our estimates are $\sqrt{n}$ consistent except $\widehat{\sigma}^2(\cdot)$, though we can get $\sqrt{n}$ consistency if we use a parametric model

The bad: The feels very ad-hoc. *Maybe* we can see this analogous to calculating the reliability ratio for regression coefficients. I think we'll have a hard time developing any theory for this method, except demonstrating consistency. I have no idea how to include uncertainty in the estimate of σ_x^2

Next?: Maybe you have a different idea for how to solve this. Maybe a likelihood based method?

References

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, 82, 1079–1091.

Ruppert, D., Wand, M. P., Holst, U., and Hossjer, O. (1998). Local Polynomial Variance Function Estimation. *Technometrics*, 39, 262–273.