eAppendix

In this appendix, we provide a concise introduction to partial least squares (PLS) regression and its relation to ordinary least squares (OLS) and principal component analysis (PCA). The citation numbers correspond to those in the reference list for the main paper.

1. Relationships between PLS, PCA and OLS regression

Partial least squares (PLS) analysis[17-21] can be viewed as an extension of principal component analysis (PCA)[26,27] and is a member of the family of continuum regression, which includes PCA and OLS.[28] From a statistical viewpoint, OLS regression is to maximise the covariance between the $n \times 1$ outcome variable vector $\mathbf{y}$ and the vector of the linear combination of covariates $\mathbf{X\beta}$, where $\mathbf{X}$ is the $n \times p$ matrix containing p covariates and $\boldsymbol{\beta}$ is the $p \times 1$ vector of OLS regression coefficients. PCA seeks to maximise the variance of principal component $\mathbf{t}_1 = \mathbf{X}\mathbf{c}_1$ under the constraint for the modulus of the $p \times 1$ vector $\mathbf{c}_1$ to be unity. Successive principal components, $\mathbf{t}_2 = \mathbf{X}\mathbf{c}_2$, $\mathbf{t}_3 = \mathbf{X}\mathbf{c}_3$ etc., are obtained by repeating the procedure on the residuals from the preceding step, and all new principal components are uncorrelated with preceding ones. The extraction of principal components uses a mathematical technique known, in matrix algebra, as singular value decomposition, which requires the calculation of eigenvectors and eigenvalues.[22,26] For p variables, $x_1, x_2,\ldots, x_p$, each principal component, pc_i, is a weighted composite of p covariates:

$$pc_i = w_{i1}x_1 + w_{i2}x_2 + \ldots + w_{ip}x_p, \qquad \text{[Eq.1]}$$

where w_{ij}, j=1 to p, is the weight for covariate x_p in principle component pc_i. For six variables without perfect multicollinearity, six principal components, which are weighted combinations of the original six covariates, can be extracted. Note that in the construction of principal components, variables $x_1, x_2,\ldots, x_p$ are usually in standardised form, i.e. they have zero means and standard deviations equal to 1. The extracted principal components are ordered by the amount of total variance across the covariates that is explained by the components, i.e. pc_1 explains more variance

than pc_2, and pc_2 explains more than pc_3, etc. The first few principal components that explain most of the covariate variance are used as revised covariates for OLS regression. If all six principal components were selected as covariates, the results, such as regression coefficients and R^2, from PCA regression would be equivalent to those from OLS regression. When the outcome is only regressed on the first few components, results will be different and this is why PCA (and also PLS and ridge regression) is a shrinkage method (although the PCA or PLS regression coefficients do not necessarily "shrink").[22] Note that the extraction of principal components does not take into account the relationship of the outcome with any of the covariates. In extreme cases, whilst the first few principal components might explain most of the variance amongst the covariates, they may have very small associations with the outcome.[27]

In contrast to PCA, PLS regression seeks to select components that maximise the covariance matrix between **y** and **t**. PLS extracts components that are also weighted combinations of the original p variables, but also takes into account their correlations with the outcome. In other words, in PCA, the extraction of components is independent of the outcome variables, whereas in PLS, components are extracted for their associations with the outcome. The extraction of PLS components is under the same constraints as with PCA: (1) the sum of the squared weights is unity and (2) the correlations amongst all components are zero.[22] When there are six covariates without perfect multicollinearity, six PLS components can be extracted (and they are independent of each other). PLS components are ordered according to the amount of variance in the outcome that is explained, i.e. the first PLS component has a higher correlation with the outcome than the second PLS component, and the second has a higher correlation than the third, etc. In PLS, the first PLS component explains most of the outcome variance as shown in our study. For p variables, x_1, x_2,…, x_p, each PLS component, $plsc_i$, is also a weighted composite of p covariates:

$$plsc_i = w_{i1}x_1 + w_{i2}x_2 + ... + w_{ip}x_p . \qquad \text{[Eq.2]}$$

The PLS regression coefficient for each x is derived from the sum of products of the regression coefficients for PLS components and the weight for each x. For example, when the outcome y is regressed on the first two PLS components, the equation is given as:

$$y = \beta_1 * plsc_1 + \beta_2 * plsc_2 + \varepsilon$$

$$= \beta_1(w_{11}x_1 + w_{12}x_2 + ... + w_{1p}x_p) + \beta_2(w_{21}x_1 + w_{22}x_2 + ... + w_{2p}x_p) + \varepsilon,$$

where β_1 and β_2 are the regression coefficients for PLS component 1 and 2, respectively, and ε is the residual error term. The PLS regression coefficient for x_1 is therefore $\beta_1 w_{11} + \beta_2 w_{21}$.

If all six PLS components are used as new covariates, the results from the PLS regression, such as regression coefficients and R^2, are equivalent to those from PCA regression and OLS regression. The advantage of PLS over PCA is that the first few components explain most of the covariance between the outcome and covariates, and as a result, the caveat of PCA regression previously discussed does not occur in PLS regression. PLS can be viewed as a middle ground between OLS regression and PCA regression.[28] When covariates are highly correlated, results from PLS will be closer to those from PCA regression, and when covariates are less correlated, results from PLS will be closer to those from OLS regression.

2. PLS and perfect collinearity

It is not feasible to estimate the 'independent' contributions of birth size, growth in body size at different stages of the lifecourse, and current body size simultaneously using OLS multiple regression, because of perfect collinearity amongst these three variables caused by their mathematical relationship. From a statistical viewpoint, this is because these three variables only have two degrees of freedom, i.e. their variable space is only 2-dimensional rather than 3-dimensional. As a result, it is not feasible to undertake OLS regression analysis, as the estimation of OLS regression coefficients involves the inversion of data matrix that contain three variables,

and a matrix with insufficient degrees of freedom (in mathematical jargon the rank of the matrix) is not invertible. However, this is not a problem for either PCA or PLS regression.

Suppose we wish to estimate the effects of zwt_0, zwt_{19} and zwt_{19-0}, for instance. Since these three covariates are perfectly collinear, at least one of them has to be omitted for OLS regression as shown in Table 2 in the main paper. In PCA and PLS, perfect collinearity amongst the three variables means that only two components can be extracted from the three covariates, but each of the two components is *a combination of the original three variables*. The first PLS component will have the largest covariance with the outcome, whereas the first PCA component has the largest explained variance amongst the covariates. As we show in the main paper, the PLS model with one component is a good approximation to the model with two components. It is important to note that results from PCA and PLS regression differ from those for OLS regression because regression coefficients for all three original variables can be estimated in PCA and PLS, whilst only regression coefficients for two out of the three variables can be estimated in OLS regression. Although there are mathematical relationships amongst the three variables, it is not possible to derive the PLS regression coefficients for the three variables from the two OLS regression coefficients. In contrast, once we obtain the three regression coefficients from PLS regression with two components, it is possible to derive the OLS regression coefficients for models with any two of the three variables, as the projection vector of the outcome variable is the same in OLS, PCA and PLS regression. The weights for each covariate in PLS components can be viewed as their share of the explained variance in the outcome, so even if the covariates are perfectly collinear, PLS is still able to partition the variance shares according to the correlations amongst the covariates and outcome. In other words, whilst OLS regression is unable to disentangle the individual contribution to the outcome of birth size, growth and current size, PLS can unravel their individual contributions according to the correlations amongst them and the outcome, by distributing the overall contribution of three variables.

From a statistical viewpoint , a model with three perfectly collinear variables and one with only two variables has a different variable space in the estimation of PCA and PLS regression coefficients. Let us use a simple example for illustration: suppose we have two standardised variables x_1 and x_2 and they are orthogonal, i.e. the correlation between them is zero, and the covariance matrix for x_1 and x_2 is:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For these two orthogonal variables, the density plot for their joint distribution is a circle with radius = 1. The radius has the same length in every direction. Now we include a third variable $x_3 = x_1 + x_2$ in the analysis, and as they are all standardised variables, their covariance matrix is:

$$\begin{bmatrix} 1 & 0 & 0.707 \\ 0 & 1 & 0.707 \\ 0.707 & 0.707 & 1 \end{bmatrix}$$

The density plot for their joint distribution is now an *ellipse*: the long axis is in the same direction as x_3 with a length of $\sqrt{2}$ (i.e. its variance is 2), and the second component (which has to be orthogonal to x_3) will be in the same direction as $x_1 - x_2$ with a length of 1. PLS analysis is first to find a vector in this variable space that has the greatest covariance with the outcome, and then to find the second vector, which is orthogonal to the first. As the variable space is different for the two matrices, the results from PLS analysis will be different. Mathematically speaking, this is because the results of singular value decomposition (SVD) on a 2 by 2 matrix (x_1 and x_2) are different from those of SVD on a 3 by 3 matrix (x_1, x_2, and x_3) with rank = 2, i.e. the density functions for the joint distribution are different. This example illustrates why and how PLS can estimate the individual contribution of the three prefect collinear variables. To gain more detail about how PLS works, the most intuitive way to explain this difference is to use vector geometry as shown in a paper by Phatak and de Jong.[22]

eTable 1: Results from partial least squares regression with 6 original body weight z-scores for 960 boys. The outcome variables are systolic (*SBP*) or diastolic blood pressure (*DBP*) at age 19 yr. The number of components means how many PLS components were extracted as covariates in the regression analysis. CumR^2 is the cumulative R^2 explained by the number of components.

Components	**1**	**2**	**6**
SBP	**Coefficient 95%CI**	**Coefficient 95%CI**	**Coefficient 95%CI**
zwt_0	-0.01 (-0.22 to 0.18)	-0.81 (-1.36 to -0.23)	-0.55 (-1.15 to 0.11)
zwt_1	0.26 (0.05 to 0.42)	-0.64 (-1.13 to -0.23)	-0.76 (-2.01 to 0.38)
zwt_2	0.37 (0.20 to 0.53)	-0.30 (-0.73 to 0.11)	0.18 (-1.11 to 1.39)
zwt_8	0.73 (0.56 to 0.90)	0.65 (0.22 to 0.94)	-0.56 (-2.18 to 0.62)
zwt_{15}	0.88 (0.68 to 1.11)	1.32 (0.91 to 1.75)	0.01 (-1.44 to 1.33)
zwt_{19}	1.06 (0.85 to 1.34)	2.10 (1.65 to 2.68)	4.19 (2.90 to 5.52)
CumR^2	7.60%	10.40%	11.56%
DBP	**Coef 95%CI**	**Coef 95%CI**	**Coef 95%CI**
zwt_0	-0.02 (-0.17 to 0.14)	-0.75 (-1.22 to -0.30)	-0.72 (-1.27 to -0.18)
zwt_1	0.33 (0.20 to 0.47)	-0.11 (-0.52 to 0.23)	-0.21 (-1.23 to 0.77)
zwt_2	0.41 (0.28 to 0.56)	0.13 (-0.26 to 0.47)	0.45 (-0.62 to 1.55)
zwt_8	0.62 (0.46 to 0.82)	0.66 (0.28 to 1.04)	0.11 (-1.15 to 1.46)
zwt_{15}	0.67 (0.51 to 0.85)	0.95 (0.63 to 1.28)	0.35 (-0.80 to 1.48)
zwt_{19}	0.74 (0.58 to 0.92)	1.27 (0.94 to 1.68)	2.20 (0.97 to 3.27)
CumR^2	6.66%	8.03%	8.33%

eTable 2: Results from partial least squares regression with 6 original body weight z-scores and 15 changes in weight z-scores for 960 boys. The outcome variables are systolic (*SBP*) or diastolic blood pressure (*DBP*) at age 19 yr. The number of components means how many PLS components were extracted as covariates in the regression analysis. CumR^2 is the cumulative R^2 explained by the number of components.

	SBP			***DBP***		
Components	**1**	**2**	**6**	**1**	**2**	**6**
	Coef 95%CI	**Coef 95%CI**	**Coef 95%CI**	**Coef 95%CI**	**Coef 95%CI**	**Coef 95%CI**
zwt_0	0.00 (-0.10 to 0.10)	0.25 (0.04 to 0.48)	0.23 (0.03 to 0.46)	-0.01 (-0.09 to 0.08)	0.13 (-0.03 to 0.33)	0.16 (-0.01 to 0.35)
zwt_{1-0}	0.11 (0.01 to 0.20)	0.07 (-0.10 to 0.22)	0.04 (-0.12 to 0.19)	0.16 (0.08 to 0.24)	0.16 (0.03 to 0.28)	0.13 (-0.01 to 0.26)
zwt_1	0.13 (0.02 to 0.23)	0.33 (0.15 to 0.49)	0.28 (0.09 to 0.45)	0.18 (0.10 to 0.28)	0.32 (0.19 to 0.46)	0.30 (0.15 to 0.45)
zwt_{2-1}	0.18 (0.00 to 0.37)	0.23 (-0.29 to 0.75)	0.28 (-0.37 to 0.89)	0.14 (-0.02 to 0.31)	0.16 (-0.21 to 0.53)	0.19 (-0.34 to 0.73)
zwt_{2-0}	0.15 (0.06 to 0.23)	0.12 (-0.03 to 0.26)	0.11 (-0.03 to 0.24)	0.18 (0.12 to 0.25)	0.19 (0.08 to 0.29)	0.17 (0.04 to 0.27)
zwt_2	0.18 (0.08 to 0.29)	0.40 (0.22 to 0.56)	0.36 (0.19 to 0.54)	0.22 (0.14 to 0.32)	0.36 (0.23 to 0.51)	0.36 (0.22 to 0.52)
zwt_{8-2}	0.24 (0.10 to 0.35)	0.03 (-0.22 to 0.24)	0.02 (-0.28 to 0.27)	0.15 (0.04 to 0.26)	0.03 (-0.18 to 0.23)	0.02 (-0.24 to 0.29)
zwt_{8-1}	0.28 (0.17 to 0.37)	0.12 (-0.12 to 0.30)	0.13 (-0.11 to 0.34)	0.18 (0.08 to 0.28)	0.09 (-0.09 to 0.26)	0.09 (-0.10 to 0.31)
zwt_{8-0}	0.24 (0.17 to 0.31)	0.12 (0.00 to 0.24)	0.11 (-0.01 to 0.22)	0.23 (0.16 to 0.31)	0.19 (0.07 to 0.30)	0.15 (0.04 to 0.28)
zwt_8	0.36 (0.27 to 0.47)	0.44 (0.27 to 0.60)	0.39 (0.23 to 0.56)	0.33 (0.23 to 0.47)	0.40 (0.25 to 0.61)	0.39 (0.23 to 0.62)
zwt_{15-8}	0.25 (0.03 to 0.50)	0.29 (-0.13 to 0.77)	0.26 (-0.22 to 0.81)	0.11 (-0.06 to 0.30)	0.10 (-0.24 to 0.46)	0.12 (-0.33 to 0.55)
zwt_{15-2}	0.27 (0.17 to 0.36)	0.13 (-0.04 to 0.28)	0.11 (-0.05 to 0.27)	0.15 (0.06 to 0.22)	0.06 (-0.08 to 0.17)	0.06 (-0.10 to 0.20)
zwt_{15-1}	0.30 (0.21 to 0.38)	0.19 (0.05 to 0.33)	0.19 (0.03 to 0.34)	0.18 (0.10 to 0.26)	0.11 (-0.01 to 0.22)	0.11 (-0.02 to 0.26)
zwt_{15-0}	0.28 (0.21 to 0.36)	0.18 (0.06 to 0.30)	0.16 (0.04 to 0.27)	0.24 (0.18 to 0.31)	0.2 (0.10 to 0.28)	0.17 (0.08 to 0.26)

zwt_{15}	0.43 (0.33 to 0.56)	0.53 (0.36 to 0.72)	0.47 (0.30 to 0.67)	0.36 (0.28 to 0.47)	0.43 (0.30 to 0.57)	0.42 (0.27 to 0.57)
$zwt_{19–15}$	0.36 (0.15 to 0.59)	0.97 (0.38 to 1.56)	1.19 (0.50 to 1.88)	0.16 (-0.04 to 0.37)	0.35 (-0.06 to 0.85)	0.53 (-0.08 to 1.10)
$zwt_{19–8}$	0.43 (0.22 to 0.66)	0.85 (0.48 to 1.30)	0.96 (0.59 to 1.43)	0.19 (0.02 to 0.36)	0.31 (-0.03 to 0.65)	0.43 (0.00 to 0.79)
$zwt_{19–2}$	0.35 (0.25 to 0.46)	0.39 (0.22 to 0.59)	0.44 (0.25 to 0.62)	0.19 (0.10 to 0.26)	0.16 (0.04 to 0.28)	0.20 (0.04 to 0.34)
$zwt_{19–1}$	0.37 (0.28 to 0.47)	0.42 (0.28 to 0.58)	0.48 (0.33 to 0.62)	0.21 (0.13 to 0.29)	0.19 (0.09 to 0.30)	0.24 (0.10 to 0.36)
$zwt_{19–0}$	0.34 (0.26 to 0.43)	0.34 (0.23 to 0.46)	0.36 (0.24 to 0.48)	0.26 (0.20 to 0.34)	0.25 (0.16 to 0.34)	0.26 (0.16 to 0.35)
zwt_{19}	0.52 (0.42 to 0.67)	0.77 (0.59 to 0.97)	0.78 (0.59 to 0.98)	0.40 (0.30 to 0.51)	0.51 (0.38 to 0.68)	0.55 (0.40 to 0.71)
CumR^2	10.23%	11.47%	11.57%	7.71%	8.24%	8.33%