Supplement C: Data Generation with the 5th Order Power Polynomial

The 5th-order power polynomial method (Headrick, 2002) was used to generate *X* and *Y* variables with specified correlations, skewness, and kurtosis. *X* and *Y* each consisted of polynomials constructed from standard normal variates (*Z and Z'*):

$$X = c_o + c_1 Z + c_2 Z^2 + \cdots + c_5 Z^5 \quad \text{(S1)}$$

$$Y = c'_o + c'_1 Z' + c'_2 Z'^2 + \cdots + c'_5 Z'^5 \ . \quad \text{(S2)}$$

In order to assure a correlation between *Z* and *Z'*, and therefore a correlation between *X* and *Y*, the normal variate *Z*' was constructed as a linear combination of *Z* and an independently generated standard normal variate, Z_0.

$$Z' = tZ + \sqrt{1 - t^2} Z_0 \ , \quad \text{(S3)}$$

where *t* is an intermediate correlation coefficient which typically differs from the target population correlation, ρ.

To simulate bivariate normal data, the constants are $c_i = c'_i = 0$ for all $i \neq 1$, $c_1 = c'_1 = 1$, and $t = \rho$. These constants assure that both *X* and *Y* are normal, and more generally, that all linear combinations of them are also normal. To simulate other situations, constants must be estimated by solving a series of polynomial equations and unknowns based on the 3rd through 6th standardized moments and the desirable population correlation, ρ. First, as recommended in Headrick (2010), values for 5th and 6th standardized moments (γ_3 and γ_4) were estimated through 3rd order method Mathematica code. For situations with no skew (γ_1=0), γ_3 was already known to be 0 and required no estimation. In one case, a γ_4 value had to be adjusted through trial and error to find the smallest approximate value with a stable solution (specifically, for γ_1= 0, γ_2= -1, a stable solution was found with γ_4=130). Next, the target γ_1 and γ_2 and estimated γ_3 and γ_4 were used as input in the 5th order Mathematica code (Headrick, Sheng, & Hodis, 2007), which was used to estimate the constants shown in Table SC1.

Table SC1

Constants For the 5th-Order Power Polynomial Data Generation Method.

										Intermediate Correlation (*t*) for ρ=.5	
γ_1	γ_2	γ_3	γ_4	c_0	c_1	c_2	c_3	c_4	c_5	When *X* & *Y* have shown moments	When *X* has shown moments, and *Y*~normal
0	-1	0	130.000	0	1.351811	0	-.171467	0	.00922413	.51919820	.51241520
0	0	0	0	0	1	0	0	0	0	.50000000	.50000000
0	2	0	24.290	0	.835664	0	.052057	0	-.00000001	.50612141	.50411522
0	4	0	98.219	0	.737382	0	.080925	0	.00000001	.51486784	.51012261
0	6	0	219.863	0	.662682	0	.101891	0	.00000000	.52367439	.51634007
0	8	0	386.838	0	.600653	0	.118712	0	.00000000	.53225522	.52258136
0	10	0	597.122	0	.546682	0	.132938	0	.00000000	.54056987	.52882240
0	20	0	2249.117	0	.338718	0	.184461	0	.00000000	.57857895	.56047454
0	30	0	4823.820	0	.179290	0	.220676	0	.00000000	.61184825	.59430653
0	40	0	8265.937	0	.041453	0	.249820	0	.00000000	.64137253	.63218184
1	2	5.584	20.677	-.147206	.904758	.147200	.023863	.00000182	-.00000041	.51211961	.51211603
1	4	15.109	99.796	-.116977	.776585	.116976	.065500	.00000017	-.00000003	.51658316	.51383046
1	6	23.821	224.238	-.102889	.690336	.102888	.091164	.00000010	-.00000002	.52423746	.51876519
1	8	31.952	392.944	-.094035	.622114	.094036	.110454	-.00000013	.00000002	.53232730	.52439761
1	10	39.652	604.397	-.087718	.564261	.087718	.126209	.00000002	.00000000	.54038671	.53028549
1	20	74.162	2258.952	-.070726	.348061	.070726	.180925	.00000001	.00000000	.57803596	.56126952
1	30	104.930	4834.381	-.062431	.185712	.062431	.218254	-.00000003	.00000000	.61127901	.59490252
1	40	133.764	8276.673	-.057246	.046374	.057246	.247965	.00000002	.00000000	.64083621	.63269574

2	8	44.468	333.629	-.233358	.710436	.233353	.072266	.00000154	-.00000036	.53904439	.53924128
2	10	62.223	565.094	-.203447	.630443	.203445	.098743	.00000044	-.00000011	.54294195	.53956621
2	20	136.665	2256.827	-.149826	.378622	.149826	.168920	.00000003	.00000000	.57685505	.56472832
2	30	200.482	4843.997	-.129343	.205931	.129343	.210419	.00000002	.00000000	.60972212	.59723732
2	40	259.463	8291.717	-.117429	.061593	.117429	.242096	-.00000003	.00000000	.63929429	.63461319
3	20	170.420	2095.789	-.260482	.442217	.260482	.140478	-.00000009	.00000002	.57865945	.57893735
3	30	274.943	4767.762	-.208933	.243540	.208933	.194735	.00000005	-.00000001	.60800770	.60405018
3	40	367.694	8249.858	-.185083	.088675	.185083	.231041	-.00000001	.00000000	.63706984	.63955121
4	30	306.683	4318.326	-.334667	.307703	.334668	.160288	-.00000022	.00000005	.61218548	.63405982
4	40	444.793	7995.129	-.271450	.130891	.271450	.211277	-.00000002	.00000000	.63559382	.65383172

Notes. γ_1=population skewness, γ_2=population kurtosis. For distributions with negative skewness (not shown here), constants for positive skewness were used except that the following values were multiplied by -1: c_1, c_3, and c_5. Intermediate correlations (*t*) for ρ=0 were all 0. Constants for *Y* variables were equal to those of *X* when they had the same distribution shape ($c'_i = c_i$). When *Y* had a normal distribution, c'_1=1, and c'_i=0 for all $i \neq 1$.