Appendix

Appendix A: Variable list

Variable Name	Wave 6 Codebook Definition	Reason for including this variable in dataset
V10	Feeling of happiness	Outcome variable for inference models
V11	State of health (subjective)	Control variable / high correlation with outcome
V23	Satisfaction with your life	Outcome variable for inference models
V24	Most people can be trusted	Control variable / high correlation with outcome
V55	How much freedom of choice and control over own life	Control variable / high correlation with outcome
V56	Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?	Control variable / high correlation with outcome
V59	Satisfaction with financial situation of household	Control variable / high correlation with outcome
V140	Importance of democracy	Control variable / high correlation with outcome
V170	Secure in neighborhood	Control variable / high correlation with outcome
V172	How frequently do the following things occur in your neighborhood: Alcohol consumed in the streets	Control variable / high correlation with outcome
V175	How frequently do the following things occur in your neighborhood: Drug sale in streets	Control variable / high correlation with outcome
V181	Worries: Losing my job or not finding a job	Control variable / high correlation with outcome
V188	In the last 12 month, how often have you or your family: Gone without enough food to eat	Control variable / high correlation with outcome
V190	In the last 12 month, how often have you or your family: Gone without needed medicine or treatment that you needed	Control variable / high correlation with outcome
V191	In the last 12 month, how often have you or your family: Gone without a cash income	Control variable / high correlation with outcome
V192	Science and technology are making our lives healthier, easier, and more comfortable	Control variable / high correlation with outcome
V237	Family savings during past year	Control variable / high correlation with outcome
V238	Social class (subjective)	Control variable / high correlation with outcome

V239	Scale of incomes	Control variable / high correlation with outcome
V240	Sex	Control variable
V242	Age	Control variable
V70	Schwartz: It is important to this person to think up new ideas and be creative; to do things one's own way	Individualistic conservatism variable for inference model
V71	Schwartz: It is important to this person to be rich; to have a lot of money and expensive things	individualistic conservatism variable for inference model
V72	Schwartz: Living in secure surroundings is important to this person; to avoid anything that might be dangerous	individualistic conservatism variable for inference model
V73	Schwartz: It is important to this person to have a good time; to "spoil" oneself	individualistic conservatism variable for inference model
V75	Schwartz: Being very successful is important to this person; to have people recognize one's achievements	individualistic conservatism variable for inference model
V76	Schwartz: Adventure and taking risks are important to this person; to have an exciting life	individualistic conservatism variable for inference model
V77	Schwartz: It is important to this person to always behave properly; to avoid doing anything people would say is wrong	individualistic conservatism variable for inference model
V78	Schwartz: Looking after the environment is important to this person; to care for nature and save life resources 1.- Very much	individualistic conservatism variable for inference model
V79	Schwartz: Tradition is important to this person; to follow the customs handed down by one's religion or family	individualistic conservatism variable for inference model

Appendix B: Proportion of missing data (PMI) of variables in our dataset

Variable	PMI	Variable	PMI
V10	0.01	V73	0.09
V11	0.01	V75	0.090
V23	0.01	V76	0.11
V24	0.04	V77	0.07
V55	0.02	V78	0.07
V56	0.03	V79	0.07
V140	0.04		
V170	0.05		
V172	0.05		
V175	0.16		
V181	0.10		
V188	0.04		
V190	0.04		
V191	0.05		
V192	0.06		
V237	0.05		
V238	0.04		
V239	0.04		
V240	0.00		
V242	0.00		
V70	0.09		
V71	0.10		
V72	0.06		

Appendix C: Covariance Coverage of variables of our dataset

	V10	V11	V23	V24	V55	V56	V59	V140	V170	V172	V175

V10	0.99	0.98	0.98	0.95	0.97	0.96	0.98	0.95	0.94	0.94	0.83
V11	0.98	1.00	0.99	0.96	0.97	0.97	0.99	0.96	0.95	0.95	0.84
V23	0.98	0.99	0.99	0.95	0.97	0.96	0.98	0.96	0.95	0.95	0.83
V24	0.95	0.96	0.95	0.96	0.94	0.94	0.95	0.93	0.92	0.92	0.81
V55	0.97	0.97	0.97	0.94	0.98	0.96	0.97	0.95	0.94	0.94	0.83
V56	0.96	0.97	0.96	0.94	0.96	0.97	0.96	0.94	0.93	0.93	0.82
V59	0.98	0.99	0.98	0.95	0.97	0.96	0.99	0.96	0.95	0.95	0.83
V140	0.95	0.96	0.96	0.93	0.95	0.94	0.96	0.96	0.93	0.93	0.82
V170	0.94	0.95	0.95	0.92	0.94	0.93	0.95	0.93	0.96	0.93	0.82
V172	0.94	0.95	0.95	0.92	0.94	0.93	0.95	0.93	0.93	0.95	0.83
V175	0.83	0.84	0.83	0.81	0.83	0.82	0.83	0.82	0.82	0.83	0.84
V181	0.90	0.90	0.90	0.88	0.89	0.89	0.90	0.88	0.88	0.88	0.78
V188	0.95	0.96	0.95	0.93	0.94	0.94	0.96	0.94	0.93	0.94	0.82
V190	0.95	0.95	0.95	0.92	0.94	0.93	0.95	0.93	0.93	0.93	0.82
V191	0.94	0.95	0.95	0.92	0.94	0.93	0.95	0.93	0.93	0.93	0.82
V192	0.93	0.94	0.93	0.90	0.93	0.92	0.93	0.92	0.91	0.91	0.81
V237	0.94	0.94	0.94	0.91	0.93	0.92	0.94	0.91	0.91	0.91	0.80
V238	0.95	0.95	0.95	0.92	0.94	0.93	0.95	0.93	0.92	0.92	0.81
V239	0.95	0.96	0.95	0.92	0.94	0.94	0.95	0.93	0.92	0.92	0.81
V240	0.99	1.00	0.99	0.96	0.98	0.97	0.99	0.96	0.95	0.95	0.84
V242	0.98	0.99	0.99	0.96	0.97	0.97	0.99	0.96	0.95	0.95	0.84
V70	0.91	0.91	0.91	0.89	0.90	0.90	0.91	0.89	0.88	0.88	0.79
V71	0.90	0.90	0.90	0.88	0.89	0.89	0.90	0.88	0.88	0.88	0.79
V72	0.93	0.93	0.93	0.91	0.92	0.92	0.93	0.91	0.91	0.90	0.80
V73	0.90	0.91	0.91	0.88	0.90	0.90	0.91	0.89	0.88	0.88	0.79
V75	0.90	0.91	0.91	0.88	0.90	0.89	0.91	0.89	0.88	0.88	0.79
V76	0.88	0.89	0.88	0.86	0.88	0.87	0.88	0.86	0.86	0.86	0.78
V77	0.92	0.93	0.93	0.90	0.92	0.92	0.93	0.91	0.90	0.90	0.80

V78	0.92	0.93	0.93	0.90	0.92	0.91	0.93	0.91	0.90	0.90	0.80
V79	0.92	0.93	0.93	0.91	0.92	0.92	0.93	0.91	0.90	0.90	0.80

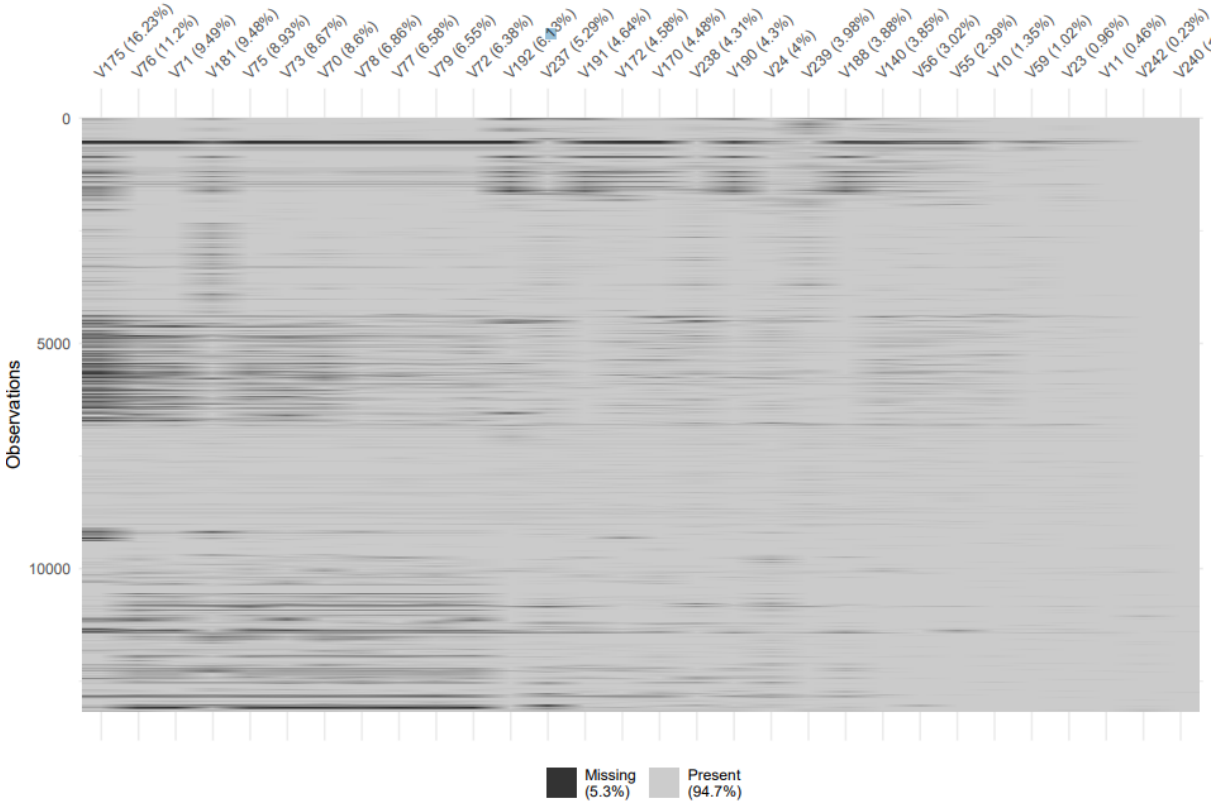
	V181	V188	V190	V191	V192	V237	V238	V239	V240	V242	V70
V10	0.90	0.95	0.95	0.94	0.93	0.94	0.95	0.95	0.99	0.98	0.91
V11	0.90	0.96	0.95	0.95	0.94	0.94	0.95	0.96	1.00	0.99	0.91
V23	0.90	0.95	0.95	0.95	0.93	0.94	0.95	0.95	0.99	0.99	0.91
V24	0.88	0.93	0.92	0.92	0.90	0.91	0.92	0.92	0.96	0.96	0.89
V55	0.89	0.94	0.94	0.94	0.93	0.93	0.94	0.94	0.98	0.97	0.90
V56	0.89	0.94	0.93	0.93	0.92	0.92	0.93	0.94	0.97	0.97	0.90
V59	0.90	0.96	0.95	0.95	0.93	0.94	0.95	0.95	0.99	0.99	0.91
V140	0.88	0.94	0.93	0.93	0.92	0.91	0.93	0.93	0.96	0.96	0.89
V170	0.88	0.93	0.93	0.93	0.91	0.91	0.92	0.92	0.95	0.95	0.88
V172	0.88	0.94	0.93	0.93	0.91	0.91	0.92	0.92	0.95	0.95	0.88
V175	0.78	0.82	0.82	0.82	0.81	0.80	0.81	0.81	0.84	0.84	0.79
V181	0.91	0.89	0.89	0.88	0.87	0.87	0.87	0.87	0.91	0.90	0.85
V188	0.89	0.96	0.95	0.95	0.93	0.92	0.92	0.93	0.96	0.96	0.89
V190	0.89	0.95	0.96	0.95	0.93	0.91	0.92	0.92	0.96	0.96	0.88
V191	0.88	0.95	0.95	0.95	0.92	0.91	0.92	0.92	0.95	0.95	0.88
V192	0.87	0.93	0.93	0.92	0.94	0.90	0.91	0.91	0.94	0.94	0.87
V237	0.87	0.92	0.91	0.91	0.90	0.95	0.92	0.92	0.95	0.95	0.87
V238	0.87	0.92	0.92	0.92	0.91	0.92	0.96	0.93	0.96	0.96	0.88
V239	0.87	0.93	0.92	0.92	0.91	0.92	0.93	0.96	0.96	0.96	0.88
V240	0.91	0.96	0.96	0.95	0.94	0.95	0.96	0.96	1.00	1.00	0.91
V242	0.90	0.96	0.96	0.95	0.94	0.95	0.96	0.96	1.00	1.00	0.91
V70	0.85	0.89	0.88	0.88	0.87	0.87	0.88	0.88	0.91	0.91	0.91
V71	0.84	0.88	0.88	0.87	0.86	0.87	0.87	0.87	0.90	0.90	0.88
V72	0.86	0.91	0.91	0.90	0.89	0.89	0.90	0.90	0.94	0.93	0.90
V73	0.84	0.89	0.88	0.88	0.87	0.87	0.88	0.88	0.91	0.91	0.88

V75	0.85	0.88	0.88	0.88	0.87	0.87	0.88	0.88	0.91	0.91	0.88
V76	0.83	0.86	0.86	0.86	0.85	0.85	0.86	0.85	0.89	0.89	0.87
V77	0.86	0.91	0.90	0.90	0.89	0.89	0.90	0.90	0.93	0.93	0.89
V78	0.86	0.91	0.90	0.90	0.89	0.89	0.90	0.90	0.93	0.93	0.89
V79	0.86	0.91	0.91	0.90	0.89	0.89	0.90	0.90	0.93	0.93	0.89

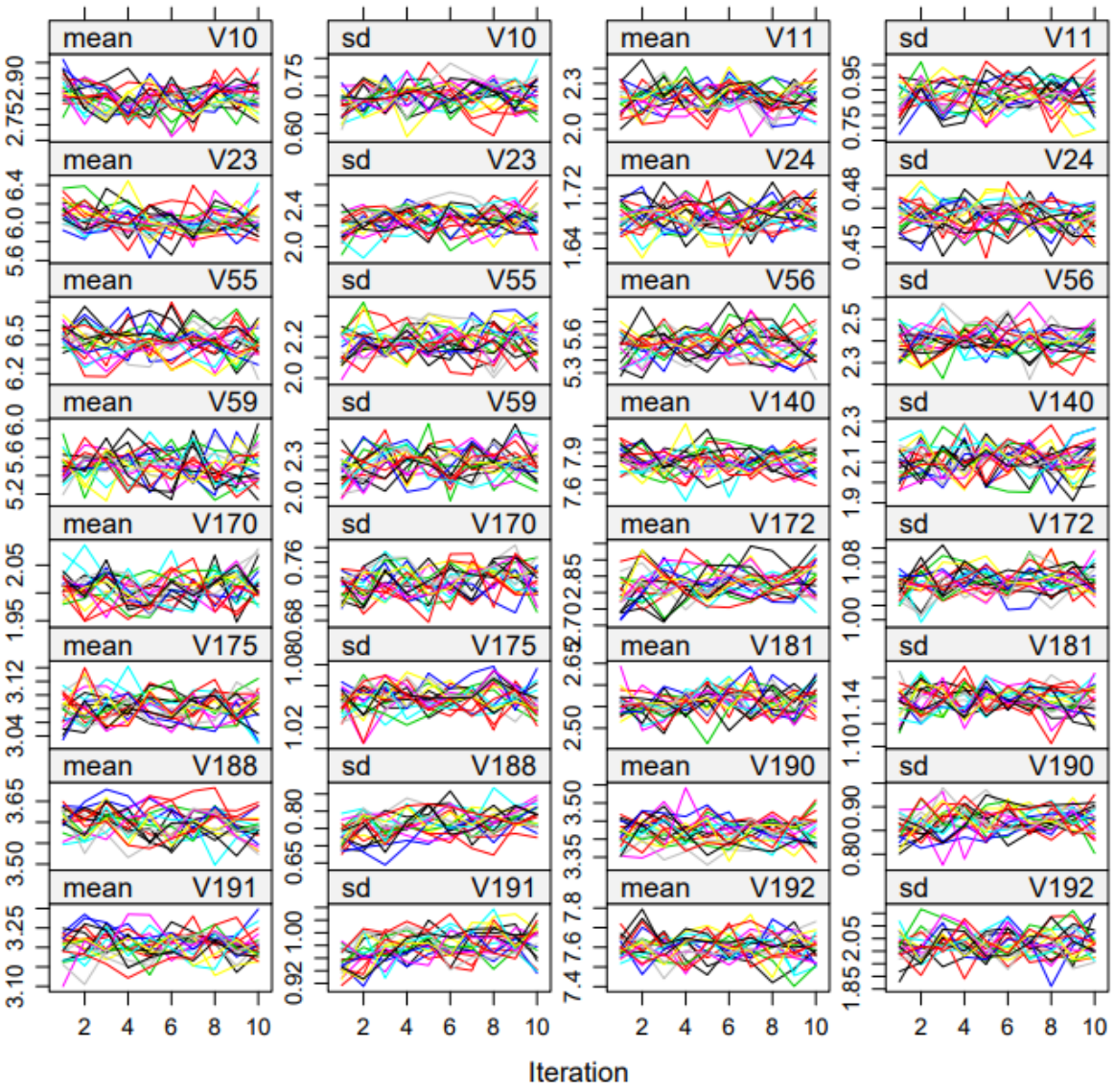
	V71	V72	V73	V75	V76	V77	V78	V79
V10	0.90	0.93	0.90	0.90	0.88	0.92	0.92	0.92
V11	0.90	0.93	0.91	0.91	0.89	0.93	0.93	0.93
V23	0.90	0.93	0.91	0.91	0.88	0.93	0.93	0.93
V24	0.88	0.91	0.88	0.88	0.86	0.90	0.90	0.91
V55	0.89	0.92	0.90	0.90	0.88	0.92	0.92	0.92
V56	0.89	0.92	0.90	0.89	0.87	0.92	0.91	0.92
V59	0.90	0.93	0.91	0.91	0.88	0.93	0.93	0.93
V140	0.88	0.91	0.89	0.89	0.86	0.91	0.91	0.91
V170	0.88	0.91	0.88	0.88	0.86	0.90	0.90	0.90
V172	0.88	0.90	0.88	0.88	0.86	0.90	0.90	0.90
V175	0.79	0.80	0.79	0.79	0.78	0.80	0.80	0.80
V181	0.84	0.86	0.84	0.85	0.83	0.86	0.86	0.86
V188	0.88	0.91	0.89	0.88	0.86	0.91	0.91	0.91
V190	0.88	0.91	0.88	0.88	0.86	0.90	0.90	0.91
V191	0.87	0.90	0.88	0.88	0.86	0.90	0.90	0.90
V192	0.86	0.89	0.87	0.87	0.85	0.89	0.89	0.89
V237	0.87	0.89	0.87	0.87	0.85	0.89	0.89	0.89
V238	0.87	0.90	0.88	0.88	0.86	0.90	0.90	0.90
V239	0.87	0.90	0.88	0.88	0.85	0.90	0.90	0.90
V240	0.90	0.94	0.91	0.91	0.89	0.93	0.93	0.93
V242	0.90	0.93	0.91	0.91	0.89	0.93	0.93	0.93
V70	0.88	0.90	0.88	0.88	0.87	0.89	0.89	0.89

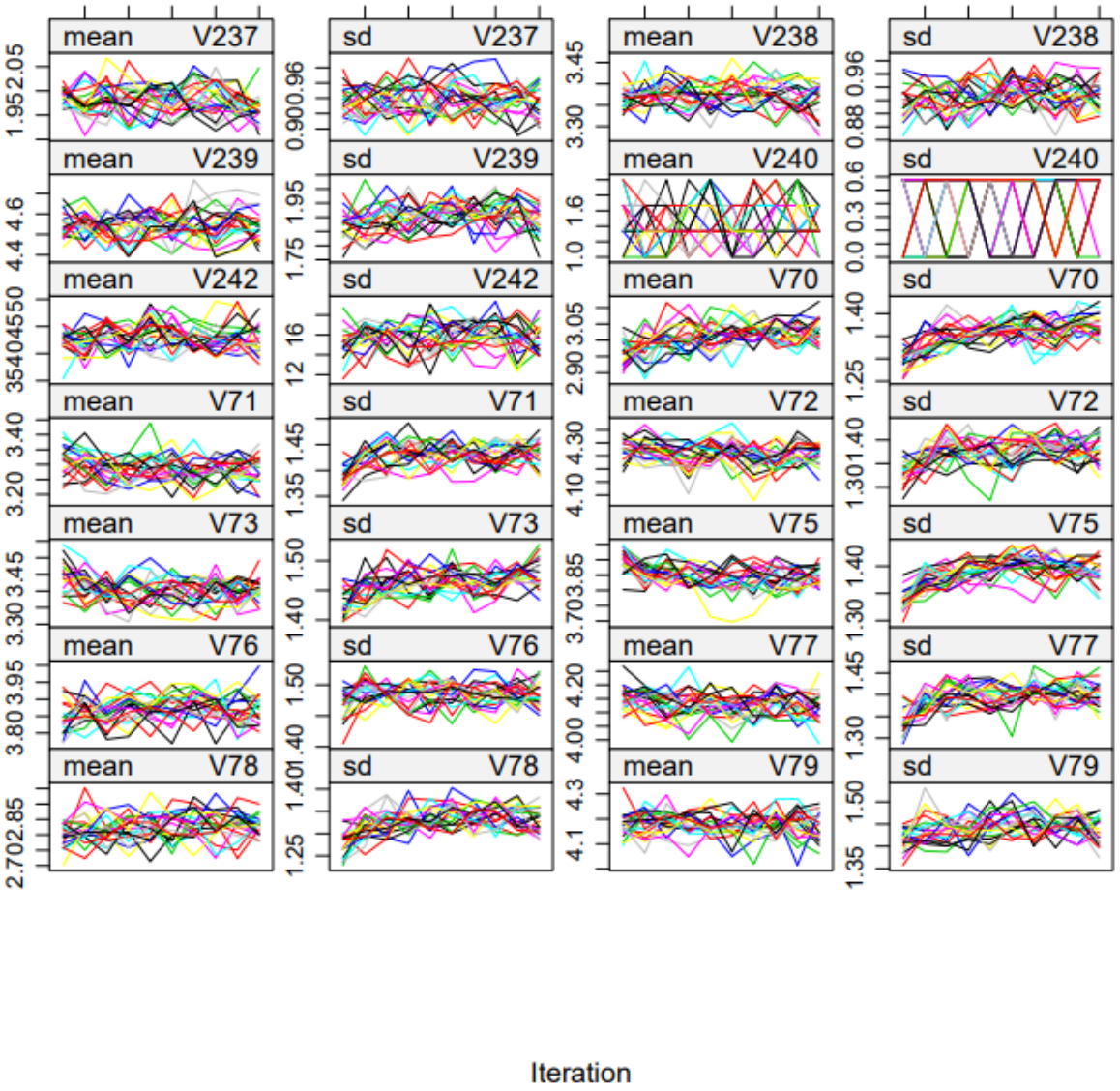
V71	0.91	0.89	0.88	0.88	0.87	0.89	0.88	0.88
V72	0.89	0.94	0.90	0.89	0.87	0.91	0.91	0.91
V73	0.88	0.90	0.91	0.88	0.87	0.89	0.89	0.89
V75	0.88	0.89	0.88	0.91	0.87	0.89	0.89	0.89
V76	0.87	0.87	0.87	0.87	0.89	0.87	0.87	0.87
V77	0.89	0.91	0.89	0.89	0.87	0.93	0.91	0.92
V78	0.88	0.91	0.89	0.89	0.87	0.91	0.93	0.91
V79	0.88	0.91	0.89	0.89	0.87	0.92	0.91	0.93

Appendix D: Missingness of data plot with vis_miss() from visdat package

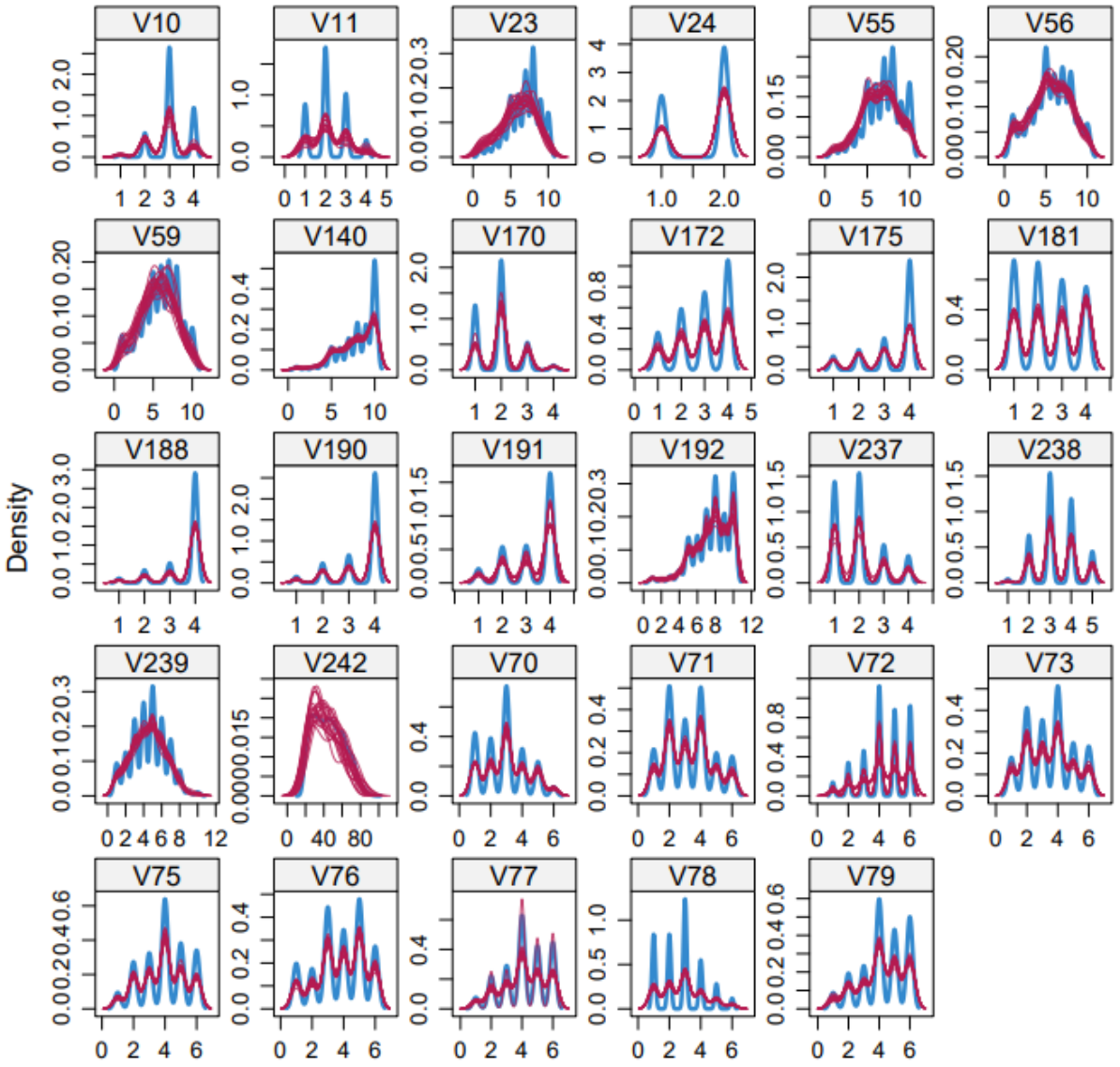


Appendix E: Convergence plot of MI datasets

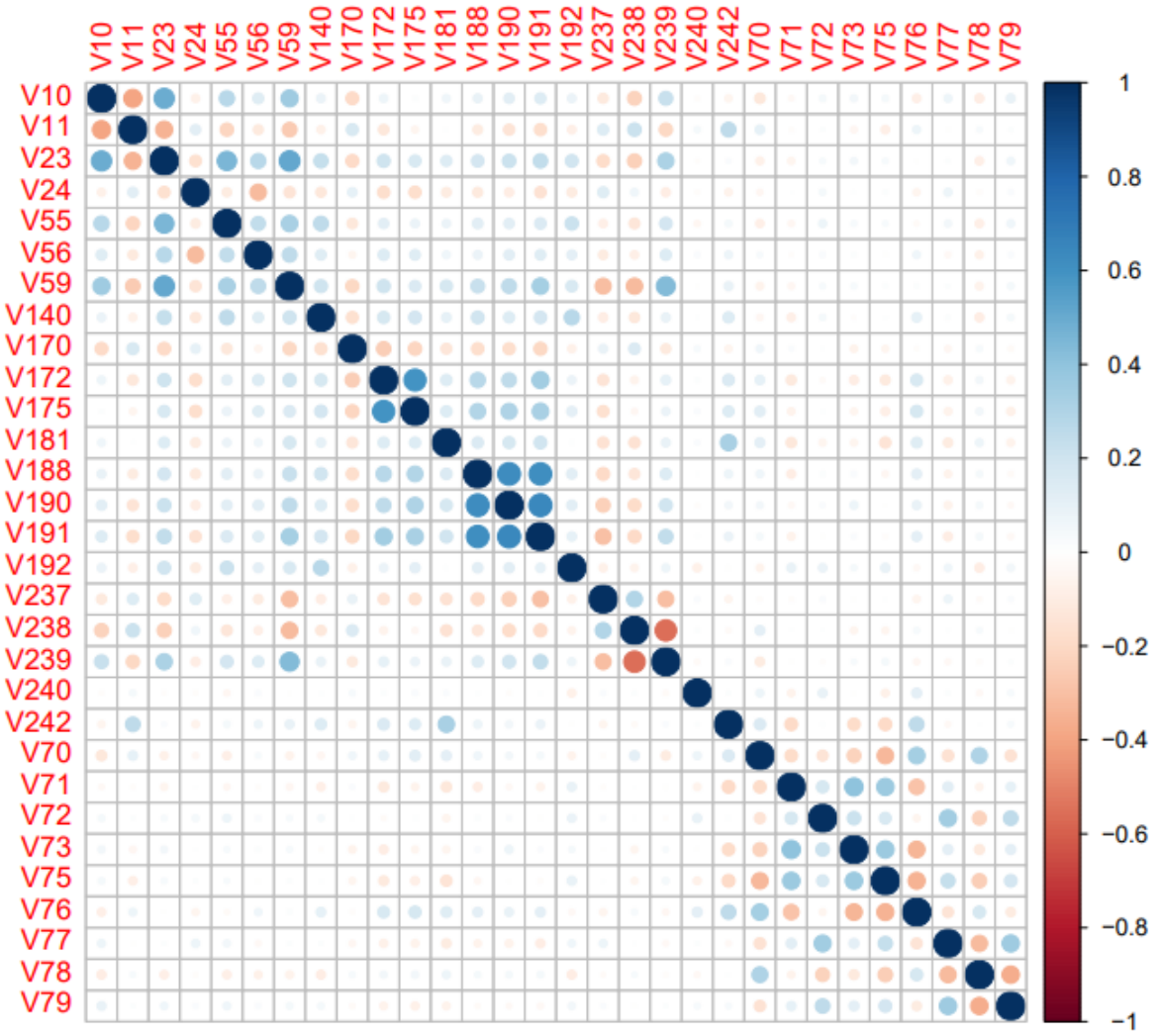



mean V237
sd V237
mean V238
sd V238
mean V239
sd V239
mean V240
sd V240
mean V242
sd V242
mean V70
sd V70
mean V71
sd V71
mean V72
sd V72
mean V73
sd V73
mean V75
sd V75
mean V76
sd V76
mean V77
sd V77
mean V78
sd V78
mean V79
sd V79
Iteration

Appendix F: Density plot of MI datasets



Appendix G: Correlation plot for Multicollinearity check



Appendix H: Correlation coefficients of variables in the dataset

Variable	Correlation coefficient (Pearson's r) with V23	Correlation coefficient (Pearson's r) With V10
V23	1.00	-0.49
V10	-0.49	1.00
V11	-0.36	0.40
V24	-0.18	0.09
V55	0.45	-0.28
V56	0.27	-0.14
V59	0.50	-0.34
V140	0.25	-0.09
V170	-0.20	0.19
V172	0.23	-0.10
V175	0.20	-0.03
V181	0.16	-0.07
V188	0.22	-0.13
V190	0.24	-0.14
V191	0.26	-0.16
V192	0.20	-0.09
V237	-0.20	0.13
V238	-0.24	0.22
V239	0.32	-0.22
V240	-0.01	0.02
V242	0.02	0.06
V70	-0.05	0.11
V71	0.07	-0.04

Variable	Correlation coefficient (Pearson's r) with V23	Correlation coefficient (Pearson's r) With V10
V72	-0.02	0.05
V73	-0.05	0.05
V75	-0.01	0.04
V76	0.02	0.07
V77	0.00	0.07
V78	-0.07	0.08
V79	-0.06	0.09

Appendix I: Control variables and reasoning

Control variable	Description	Reason for including in the inference model
V11	State of health (subjective)	Moderate/high correlation with the outcome variable V23 (r = 0.45 , t [6142] = 39.50, p < .001)
V55	How much freedom of choice and control over own life	Moderate/high correlation with the outcome variable (r = 0.50 , t[6142] = 45.05, p < .001)
V59	Satisfaction with financial situation of household	Might have an influence on life satisfaction and happiness according to theory (Ryff & Keyes, 1995) and moderate/high correlation with the outcome variable
V240	Sex	Might have an influence on life satisfaction and happiness according to theory (Ryff & Keyes, 1995)
V242	Age	Might have an influence on life satisfaction and happiness according to theory (Ryff & Keyes, 1995)
V10/V23	Feeling of happiness / Life satisfaction	Moderate/high correlation with the outcome variable (r = -0.49 , t[6142] = -44.18, p < .001)

Appendix J: Regression coefficients of inference models (without Outliers)

Table 1: Model 1a - Regression coefficients for inference model (DV: Life satisfaction ~ Schwartz variables (conservative attitude) + control variables)

Term (IV)	β	*SE*	T-statistic	*p*-value
(Intercept)	1.149	0.139	8.24	0.000
V10	0.926	0.025	37.04	0.000
V11	-0.249	0.020	-12.67	0.000
V55	0.240	0.007	32.24	0.000
V59	0.257	0.007	37.22	0.000
V70	-0.002	0.013	-0.19	0.853
V71	0.042	0.012	3.52	0.000
V72	0.012	0.012	1.04	0.300
V73	-0.028	0.012	-2.37	0.018
V75	-0.033	0.013	-2.58	0.010
V76	0.030	0.012	2.47	0.014
V77	-0.034	0.013	-2.67	0.008
V78	0.016	0.0134	1.21	0.228
V79	0.011	0.012	0.92	0.356
V240Females	0.076	0.028	2.70	0.007
V242	0.001	0.001	0.89	0.373

Table 2: Model 1b - Regression coefficients for inference model (DV: Level of Satisfaction ~ Mean scale score representing conservative attitude + control variables)

Term (IV)	β	*SE*	*T* -statistic	*p*-value
(Intercept)	1.110	0.136	8.15	0.000
Mean_scale_score	0.020	0.027	0.73	0.465
V10	0.929	0.025	37.32	0.000
V11	-0.25	0.020	-12.78	0.000
V55	0.242	0.007	32.72	0.000
V59	0.259	0.007	37.71	0.000
V240Females	0.083	0.028	2.96	0.003
V242	0.001	0.001	1.27	0.203

Table 3: Model 2a - Regression coefficients for inference model (DV: Feeling of Happiness ~ Schwartz variables (conservative attitude) + control variables)

Term (IV)	β	*SE*	T-statistic	*p*-value
(Intercept)	2.45	0.043	56.41	0.000
V23	0.111	0.003	36.8	0.000
V11	-0.193	0.007	-29.05	0.000
V55	0.011	0.003	3.96	0.000
V59	0.025	0.003	9.81	0.000
V70	-0.018	0.004	-4.08	0.000
V71	0.015	0.004	3.45	0.001
V72	0.000	0.004	0.00	0.998
V73	-0.01	0.004	-2.35	0.019
V75	0.014	0.004	3.01	0.003
V76	-0.017	0.004	-4.25	0.000
V77	0.011	0.004	2.61	0.009
V78	0.001	0.005	0.12	0.905
V79	0.024	0.004	5.86	0.000
V240Females	0.002	0.010	0.19	0.847
V242	0.000	0.000	-0.17	0.863

Table 4: Model 2b - Regression coefficients for inference model (DV: Feeling of Happiness ~ Mean scale score representing conservative attitude + control variables)

Term (IV)	β	*SE*	T-statistic	*p*-value
(Intercept)	2.475	0.043	57.93	0.000
Mean_scale_score	0.016	0.010	1.66	0.097
V23	0.113	0.003	37.10	0.000
V11	-0.192	0.007	-28.81	0.000
V55	0.013	0.003	4.73	0.000
V59	0.025	0.004	9.81	0.000
V240Females	0.000	0.010	-0.03	0.974

Appendix K: Model specification for inference

Model	DV	IV's
Model 1a	Life satisfaction (V23)	Schwartz variables (conservative attitude) + control variables. V10 + V11 + V55+ V59 + V70 + V71 +V72 + V73 + V75 + V76 + V77 + V78 + V79 + V240 + V242
Model 1b	Life satisfaction(V23)	Mean scale score representing conservative attitude + control variables. Mean_scale_score + V10 + V11 + V55+ V59 + V240 + V242
Model 2a	Feeling of happiness (V10)	Schwartz variables (conservative attitude) + control variables. V10 + V11 + V55+ V59 + V70 + V71 +V72 + V73 + V75 + V76 + V77 + V78 + V79 + V240 + V242
Model 2b	Feeling of happiness (V10)	Mean scale score representing conservative attitude + control variables. Mean_scale_score + V23 + V11 + V55+ V59 + V240 + V242

Appendix L: Prediction model: specification and evaluation

Table 1: Model specification and Test MSE

Model	Variables in model	Train CVE	Test MSE
1. Only statistically significant regression coefficients (β) ($p < 0.001$)	V23 ~ V10 + V11 + V55 + V59 + V71 + V73	2.282	2.302
2. Only statistically significant regression coefficients (β) ($p < 0.01$)	V23 ~ V10 + V11 + V55 + V59 + V71 + V73 + V75 + V76 + V77 + V240	2.278	2.299
3. Only statistically significant regression coefficients (β) ($p < 0.05$)	V23 ~ V10 + V11 + V55 + V59 + V71 + V72 + V73 + V75 + V76 + V77 + V240	2.278	2.299
4. Moderately correlated variables w/ outcome ($r >= 0.30$)	V23 ~ V10 + V11 + V55 + V59 + V239	2.266	2.291
5. Moderately correlated variables w/ outcome ($r >= 0.25$)	V23 ~ V10 + V11 + V55 + V56 + V59 + V191 + V239	2.248	2.266
6. Moderately correlated variables w/ outcome ($r >= 0.20$)	V23 ~ V10 + V11 + V24 + V55 +V56 + V59 + V140 + V170 + V172 + V175 + V181 + V188 + V190 + V191 + V192 +V237 +V238 +V239	2.195	2.210
7. Model with all variables	V23 ~ . - V23 - Mean_scale_score	2.192	2.203

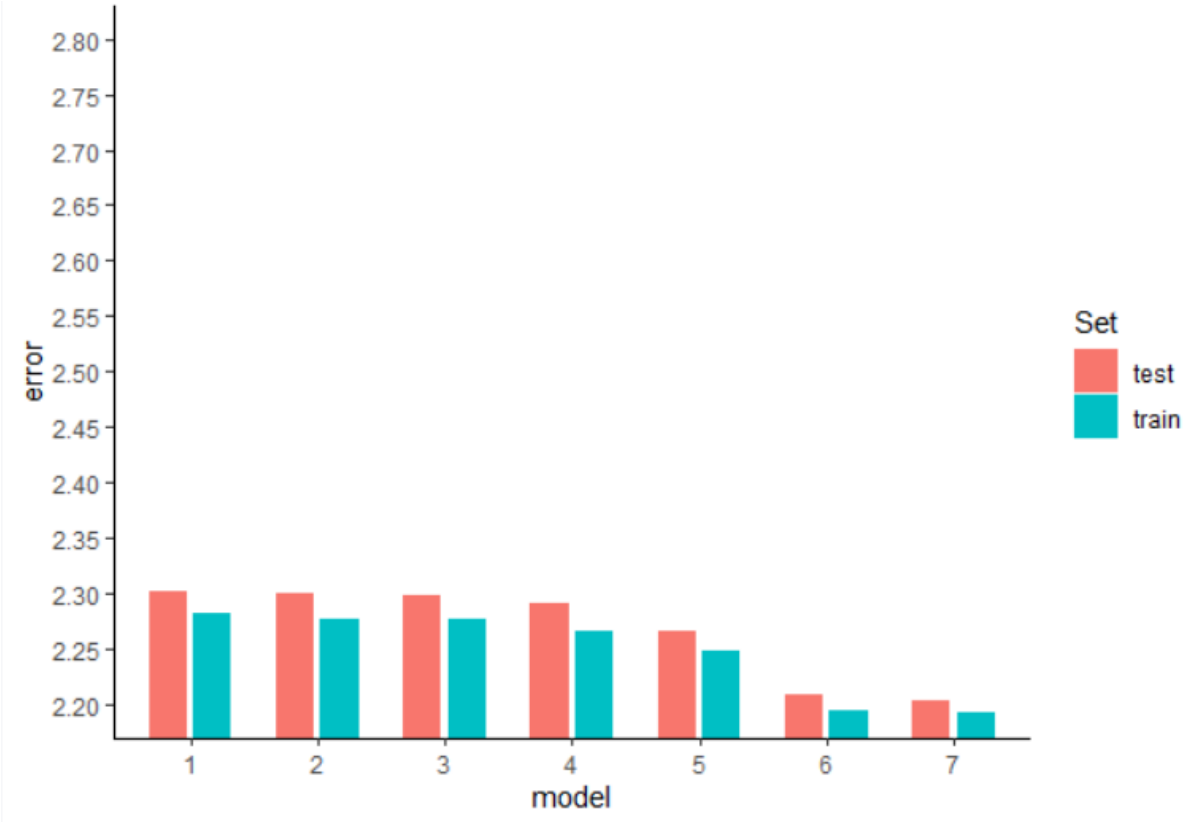


Figure 1: Test MSE (error) per model and train/test set

Appendix M: Inferential modeling including outliers

This appendix is dedicated to reporting the inferential analysis of section 2.2, on a dataset that the multivariate outliers were maintained in the dataset. I will compare the difference between the model on dataset with and without multivariate outliers

In the first model, with life satisfaction (V23) as dependent variable ($\Delta R2$ = 0.43, F[15,13140] = 613.068, p < .001, RIV =.088). This time, p-values of V72 (smallest partial effect in a model without outliers) are also statistically insignificant. In the set of variables whose regression slopes are statistically significant; V77 becomes to have the largest (partial) effect on outcome,,asides from control variables (β = -.05, SE = .012, t[13140] = 4.082, p < .001, FMI = .12). It can also be seen that the direction of all statistically significant slopes remain unchanged. A significant regression equation from Mean_scaled_score on V23 was found ($\Delta R2$ = .43, F[15,13148] = 1372.55, p < .001, RIV = .003). However, the slope for Mean_Scaled_Score is still statistically insignificant and the direction of it remains unchanged.

Similarly, in the second model, I am looking at Feeling of happiness (V10) as the outcome ($\Delta R2$ = .331, F[15,13140] = 392.18, p < .001, RIV = .085). Next, the variables with statistically insignificant regression slope in the second model are still V72, V73, V78. Meanwhile, in the set of variables whose regression slopes are statistically significant; in different with this own model in dataset without outliers, the largest effect, asides from control variables belongs to V71 (β = -.023, SE = .004, t[13140] = -5.567, p < .001, FMI = .0124). Again, I want to conclude the effect of conservative attitudes on feelings of happiness. From the regression output,again, the direction of all statistically significant slopes remain unchanged. Besides that, a significant regression equation from Mean_scaled_score on V10 was also found ($\Delta R2$ = .25, F[15,13149] = 704.47, p < .001, RIV = .04). However, this time, the slope for Mean_Scaled_Score is now statistically insignificant (β = 2.091, SE = .001, t[13140] = 1.971, p < .05, FMI = .12). This means I have enough evidence to interpret that Conservative attitude will have a good effect on Feeling of Happiness in case I remain multivariate outliers in the dataset.

Appendix N: 2.3 Predictive modeling including outliers

The same procedure of feature selection for potential models, applied a 80/20 splitting for the dataset with multivariate outliers and conducted K-fold Cross-Validation (K=10) to train all models. The final model was chosen based on the lowest cross-validation error, re-fitted to the training set and evaluated based on the test set performance.

According to the result of training and evaluating our models, model [7] (includes features with moderate Pearson's correlation coefficient $r > 0.2$, Schwartz features and control features) had the lowest cross-validation error CVE = 2.513. Therefore, I selected this model as our final model. I then re-fitted model [7] to our training partition, and used that re-fitted model to evaluate performance based on the test partition. The test-set prediction error for our final model was MSE = 2.567. The prediction performance of model [7] when modeling with multivariate outliers was not as good as when modeling without outliers, represented in higher errors (both CVE and MSE).