Supplementary Document for the Manuscript entitled "Robust Functional Principal Component Analysis Based on a New Regression Framework"

S1 Additional Simulation Studies

S1.1 Varying Measurement Error Variance

We compare the estimation precision of the methods under increasing measurement error variance by varying the values of σ_e. The default values of the simulation parameters are the same as the previous section, and we set the number of curves $n_c = 200$. We experiment with cases where the standard deviation of the measurement error term $\sigma_e = 0.01$, 0.1 and 0.5. Note that in the data generation scheme, the variance of the third FPC score is 1, and thus it is reasonable to assume that σ_e would be no larger than 1.

Table S1 shows the IMSE of the three FPCs and MSE of the three FPC scores with varying values of σ_e, averaged across 500 data replications. Compared with the PACE method, the superiority of the proposed method is demonstrated by the smaller mean squared error in the estimation of FPC and FPC scores. In general, under increasing values of σ_e, both the IMSE and MSE of the FPCs and FPC scores appear to follow an increasing trend, albeit the degree of increase is relatively small. For the first two FPCs, the increase in the IMSE and MSE is not evident; for example, when σ_e increases from 0.1 to 0.5, the IMSE of the second FPC merely changes from 0.0018 to 0.0019. This suggests that the performance of the proposed method is not highly sensitive to the measurement error variance in the estimation of the first two FPCs. For the third FPC, we observe that the degree of increase in the IMSE and MSE is quite large; for example, when σ_e increases from 0.1 to 0.5, we observe more than a one-fold increase in the IMSE and MSE of the third FPC, changing from 0.1947 to 0.4024, and from 0.2677 to 0.5739, respectively. This is possibly due to the fact that the measurement error variance becomes relatively close to the variance of the third FPC, and thus it is harder to distinguish the third FPC from the noise.

S1.2 Varying Data Contamination Mechanism

We consider various data contamination mechanisms and study the performance of our method under varying values of simulation parameters τ_c, τ_p and $\mu^{(c)}$. The parameter τ_c controls the probability that a whole curve is subject to contamination, τ_p controls the proportion of data points on a curve that are contaminated, and $\mu^{(c)}$ controls the degree of contamination. As a default, we set $n_c = 200$, $\sigma_e = 0.01$ and $\sigma_c = 0.01$, and compute the IMSE and MSE across 500 data replications.

S1.2.1 Curve-Specific Contamination

We experiment with cases where the curve-specific contamination probability $\tau_c = 0.3$, 0.4, and 0.5, with $\tau_p = 0.3$ and $\mu^{(c)} = 10$. Table S2 shows the IMSE of the three FPCs and MSE of the three FPC scores with varying values of τ_c, averaged across 500 data replications. For an increasing value of τ_c, i.e., a higher proportion of curves being subject to contamination, the estimation error increases. It is worth noting that the relative magnitude of increase in the IMSE and MSE of the FPC and FPC scores is larger for the higher order FPC. For example, when the curve-specific contamination probability τ_c is as large as 0.5, we observe that both IMSE and MSE of the third FPC of the proposed method almost double to 0.9970 and 1.6573, respectively. These suggest that the estimation of the higher order FPC and FPC scores is more challenging with an increasing degree of contamination. In all cases, the proposed method has smaller mean squared errors in the estimation of FPC and FPC scores compared with the PACE method.

It is worth noting that the improvement of the IMSE over the PACE method is more significant for the higher order FPC. The reason is that the variance associated with the higher order FPC, i.e., the variance of the FPC score α_{im}, is smaller. As a result, it would harder to distinguish the higher order FPC from the observed trajectories. The IMSE of $\hat{\psi}_m$ would be more sensitive to the data contamination. The PACE method would thus perform much poorer than the proposed method.

S1.2.2 Point-Specific Contamination

We also consider cases varying point-specific contamination probability $\tau_p = 0.3$, 0.4, and 0.5, with $\tau_c = 0.3$ and $\mu^{(c)} = 10$. Table S3 shows the results with varying values of τ_p, averaged across 500 data replications. Similar to the case with curve-specific contamination, we observe that the IMSE and MSE of the FPC and FPC scores increase with the increasing value of τ_p. Moreover, the relative magnitude of increase in the estimation error is much larger for higher order FPC. While the proposed method outperforms the PACE method in terms of the IMSE of the FPCs, there are cases where the PACE method performs slightly better than the proposed method when it comes to the estimation of FPC scores.

S1.2.3 Degree of Directional Shift

Furthermore, we examine the effect of the mean directional shift applied to a contaminated data point. We consider cases with varying values of the mean directional shift while fixing $\tau_c = \tau_p = 0.1$. Table S4 shows the results with varying values of $\mu^{(c)} = 20$, 30, and 40, averaged across 500 data replications. In these cases, we observe that the PACE method performs much worse than the proposed method when the amount of directional shift is large. For example, when $\mu^{(c)}$ increases from 20 to 40, the MSE of the third FPC increases from 0.0881 to 2.8883 under the PACE method, whereas there is almost no change of the MSE under the proposed method. It is worth noting that in this set of simulations, the τ_c and τ_p take a value of 0.1, which is smaller than the ones in other simulation sets. The idea in this set of simulation is to show that even a very small proportion of contaminated data points would lead to considerably worse estimation accuracy under the PACE method, whereas the estimation accuracy of the proposed method is quite robust to the degree of directional shift in the contaminated cases, as long as the proportion of contamination is low.

Table S1: Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of σ_e, the standard deviation of the measurement error. The values are averaged across 500 data replications, with standard error shown in parentheses.

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$\sigma_e = 0.01$						
PACE	0.0015	0.0027	1.8031	0.5353	0.2322	2.6927
	(0.0016)	(0.0021)	(0.1373)	(0.6654)	(0.1625)	(0.3103)
Robust FPCA	0.0016	0.0018	0.1936	0.5120	0.1999	0.2640
	(0.0019)	(0.0020)	(0.2430)	(0.6889)	(0.1699)	(0.3048)
$\sigma_e = 0.1$						
PACE	0.0015	0.0027	1.8061	0.5353	0.2323	2.7078
	(0.0016)	(0.0021)	(0.1362)	(0.6654)	(0.1626)	(0.3135)
Robust FPCA	0.0016	0.0018	0.1947	0.5124	0.2004	0.2677
	(0.0019)	(0.0020)	(0.2333)	(0.6884)	(0.1692)	(0.2992)
$\sigma_e = 0.5$						
PACE	0.0015	0.0029	1.8428	0.5354	0.2330	2.9692
	(0.0016)	(0.0021)	(0.1137)	(0.6654)	(0.1626)	(0.3187)
Robust FPCA	0.0016	0.0019	0.4024	0.5248	0.2122	0.5739
	(0.0019)	(0.0019)	(0.2583)	(0.6881)	(0.1666)	(0.3776)

Table S2: Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of τ_c, the probability that a curve is subject to data contamination. The values are averaged across 500 data replications, with standard error shown in parentheses.

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$\tau_c = 0.3$						
PACE	0.0015	0.0027	1.8031	0.5353	0.2322	2.6927
	(0.0016)	(0.0021)	(0.1373)	(0.6654)	(0.1625)	(0.3103)
Robust FPCA	0.0016	0.0018	0.1936	0.5120	0.1999	0.2640
	(0.0019)	(0.0020)	(0.2430)	(0.6889)	(0.1699)	(0.3048)
$\tau_c = 0.4$						
PACE	0.0015	0.0031	1.8317	0.5551	0.2555	2.9673
	(0.0017)	(0.0023)	(0.1318)	(0.6651)	(0.1658)	(0.2763)
Robust FPCA	0.0016	0.0018	0.5268	0.5464	0.2267	0.8150
	(0.0018)	(0.0019)	(0.2887)	(0.6963)	(0.1730)	(0.4198)
$\tau_c = 0.5$						
PACE	0.0016	0.0031	1.8544	0.5787	0.2757	3.0923
	(0.0017)	(0.0022)	(0.1041)	(0.6670)	(0.1658)	(0.2096)
Robust FPCA	0.0017	0.0021	0.9970	0.5743	0.2594	1.6593
	(0.0017)	(0.0019)	(0.4745)	(0.6911)	(0.1611)	(0.7046)

Table S3: Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of τ_p, the probability that the data points on a curve are affected by contamination. The values are averaged across 500 data replications, with standard error shown in parentheses.

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$\tau_p = 0.3$						
PACE	0.0015	0.0027	1.8031	0.5353	0.2322	2.6927
	(0.0016)	(0.0021)	(0.1373)	(0.6654)	(0.1625)	(0.3103)
Robust FPCA	0.0016	0.0018	0.1936	0.5120	0.1999	0.2640
	(0.0019)	(0.0020)	(0.2430)	(0.6889)	(0.1699)	(0.3048)
$\tau_p = 0.4$						
PACE	0.0016	0.0035	1.9027	0.5454	0.2439	4.1998
	(0.0016)	(0.0029)	(0.0739)	(0.6675)	(0.1597)	(0.3642)
Robust FPCA	0.0017	0.0019	0.1922	0.5750	0.2586	0.6506
	(0.0018)	(0.0019)	(0.1874)	(0.6874)	(0.1637)	(0.5309)
$\tau_p = 0.5$						
PACE	0.0017	0.0053	1.9357	0.5512	0.2600	6.0500
	(0.0016)	(0.0050)	(0.0499)	(0.6684)	(0.1608)	(0.4540)
Robust FPCA	0.0018	0.0025	0.7968	0.5844	0.2681	3.0898
	(0.0018)	(0.0022)	(0.4239)	(0.6865)	(0.1608)	(1.3348)

Table S4: Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of $\mu^{(c)}$, the mean size of the directional shift in the contamination. The values are averaged across 500 data replications, with standard error shown in parentheses.

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$\mu^{(c)} = 20$						
PACE	0.0014	0.0020	0.0213	0.5077	0.2056	0.0881
	(0.0017)	(0.0019)	(0.0212)	(0.6693)	(0.1627)	(0.0365)
Robust FPCA	0.0017	0.0021	0.0013	0.4980	0.1955	0.0202
	(0.0022)	(0.0023)	(0.0028)	(0.6886)	(0.1953)	(0.0175)
$\mu^{(c)} = 30$						
PACE	0.0015	0.0025	1.0017	0.5518	0.2544	1.3819
	(0.0017)	(0.0021)	(0.5951)	(0.6717)	(0.1643)	(0.6764)
Robust FPCA	0.0017	0.0021	0.0017	0.4977	0.1951	0.0204
	(0.0022)	(0.0023)	(0.0028)	(0.6852)	(0.1961)	(0.0178)
$\mu^{(c)} = 40$						
PACE	0.0016	0.0033	1.6072	0.6183	0.3328	2.8883
	(0.0017)	(0.0027)	(0.2957)	(0.6926)	(0.1747)	(0.5398)
Robust FPCA	0.0018	0.0021	0.0011	0.5020	0.2004	0.0196
	(0.0022)	(0.0023)	(0.0016)	(0.6987)	(0.2017)	(0.0176)

S2 Additional Results from the Kidney Glomerular Filtration Rate Data Analysis

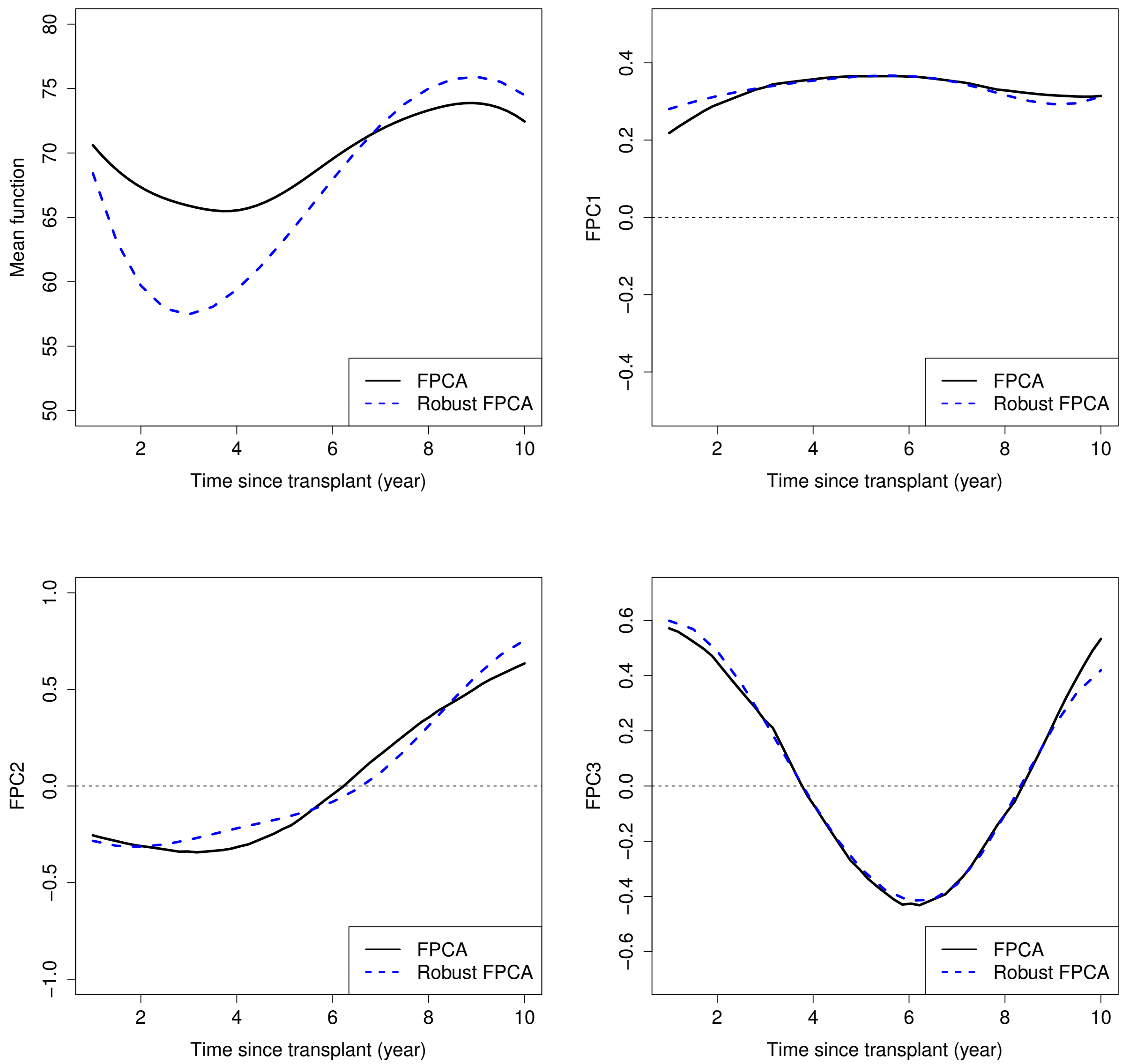

Figure S1: The estimated mean function and FPCs for the GFR trajectories in the kidney transplant data.

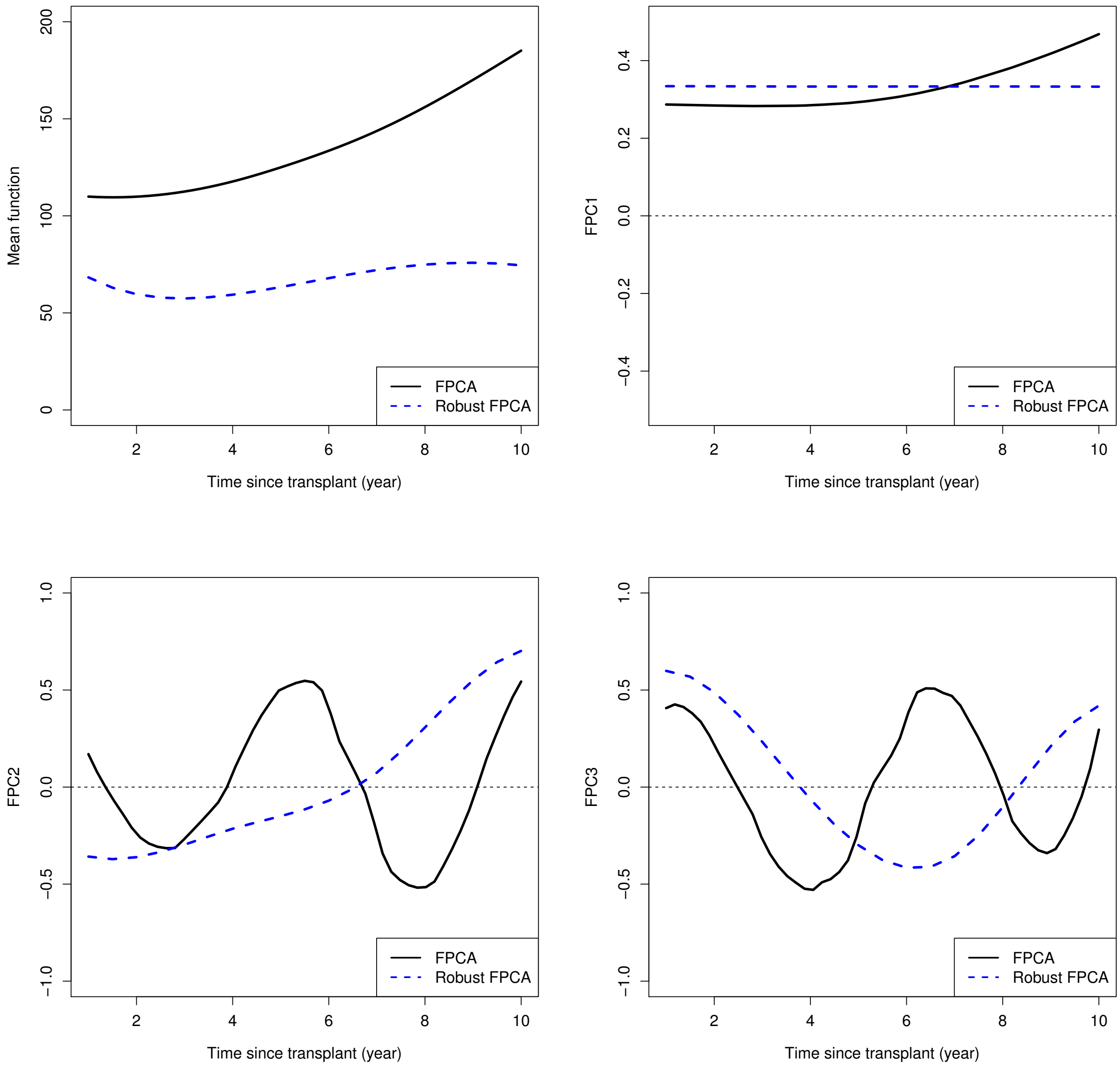

Figure S2: The estimated mean function and FPCs for the GFR trajectories in the kidney transplant data with introduced outliers.