Online Resource 2

1 Introduction

This report contains supplementary material for the paper entitled "Multiple imputation for bounded variables" (hereafter, the Manuscript). This material consists of additional results, figures, and tables concerning the simulation study and the Celtic Country Teacher Survey data analysis.

2 Supplementary tables and figures

Tables 1-5 show the simulation results described in Section 4 of the Manuscript for $n = 100$. Tables 6-10 show the simulation results described in Section 4 of the Manuscript for $n = 1000$. Figures 1-3 are supplementary figures for the analysis of the Celtic Country Teacher Survey data in Section 5 of the Manuscript.

Table 1. Average relative bias (%) at different probabilities of the empirical cumulative distribution of Y_5 ($n = 100$) for the complete case analysis (CC), normal imputation (LM), log-normal imputation (LMlog), predictive mean matching (PMM), linear quantile regression imputation (QR), and imputation based on the symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LM	LMlog	PMM	QR	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1									
0.01	-30.4	511.1	-33.3	-19.6	413.1	4.4	-7.9	98.8	65.5
0.05	-24.5	61.1	-17.2	-7.2	46.8	0.5	-7.1	24.4	12.8
0.25	-15.7	-4.3	0.8	-0.9	-4.1	0.4	-0.8	2.9	1.5
0.5	-9.7	-4.5	1.8	-0.4	-3.3	-0.1	0.4	-0.3	0.2
0.75	-4.6	-0.4	0.4	-0.1	-0.0	0.0	0.3	-0.7	-0.0
0.95	-1.1	0.1	-0.3	0.0	0.1	0.0	-0.0	-0.5	-0.2
Model 2									
0.01	15.4	479.4	8.3	15.1	199.0	18.2	14.4	20.9	19.4
0.05	5.3	83.9	7.2	10.4	33.3	7.8	5.4	10.5	8.4
0.25	-5.5	0.3	2.3	2.0	-0.6	0.8	-1.2	1.4	-0.6
0.5	-6.4	-6.3	-0.2	-0.4	-2.9	-0.5	-1.1	-0.4	-0.9
0.75	-4.4	-0.8	-1.5	-1.0	-0.3	-0.5	-0.2	-0.6	-0.1
0.95	-1.2	-0.2	-1.9	-0.3	0.0	-0.1	-0.1	-0.2	-0.1
Model 3									
0.01	2.0	106.2	-19.0	0.8	40.8	11.9	8.9	9.3	9.8
0.05	2.0	11.1	-6.8	-1.0	7.1	2.5	1.2	3.5	3.7
0.25	2.5	-7.2	12.1	0.5	1.5	0.3	-0.1	0.3	0.1
0.5	2.2	-0.2	7.2	0.3	0.2	-0.3	-0.7	-0.3	-0.9
0.75	1.7	2.9	0.8	0.7	-0.0	0.2	-0.2	0.2	-0.5
0.95	0.7	-0.7	-4.0	0.4	-0.6	-0.0	-0.3	-0.1	-0.6

Table 2. The first two columns show the pooled estimated coefficients and standard errors (SE) from the logistic regressions on $\Pr\left(I_{Y_5<0.1}\right)$ ($n = 100$) using the full datasets (FD). The remaining columns show the average relative differences (%) as compared to those from FD for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	FD		CC		LMlog		PMM		QRTs (1)		QRTa (1)		QRTs (2)		QRTa (2)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Model 1																
θ_0	3.5	1.1	4.8	28.1	2.7	22.3	−1.0	23.2	2.6	25.0	−2.2	23.4	7.2	24.9	1.0	24.6
θ_1	−2.0	0.6	3.8	20.0	0.5	15.8	−1.0	15.8	1.7	18.5	−2.1	15.7	4.2	17.8	−0.6	16.7
θ_2	−2.0	0.6	3.3	18.1	−3.3	14.3	−1.4	15.0	1.1	16.6	−3.5	14.7	1.2	16.8	−3.2	15.1
θ_3	−0.2	0.3	7.2	18.0	2.8	15.5	6.6	15.6	5.1	15.2	−0.8	14.6	9.0	17.1	3.7	15.5
Model 2																
θ_0	1.1	0.7	−12.5	24.0	−4.0	22.8	2.6	23.3	−0.5	23.4	−11.9	24.5	0.7	23.4	−10.3	23.4
θ_1	−0.6	0.4	−7.7	14.4	−3.3	13.6	2.7	14.8	0.5	15.2	−7.3	15.6	1.7	15.3	−6.0	14.8
θ_2	−0.8	0.4	−7.5	10.5	−11.5	11.2	−11.2	10.9	−1.5	12.0	−6.0	11.8	−0.4	12.1	−5.4	11.3
θ_3	−0.1	0.2	−7.9	11.1	−9.6	11.5	−12.5	11.9	−3.1	12.9	−4.2	12.1	−6.1	13.3	−6.6	11.9
Model 3																
θ_0	−1.6	0.5	1.0	51.2	−24.0	41.0	7.3	61.6	5.7	54.9	6.9	53.1	6.3	53.7	6.1	53.0
θ_1	0.2	0.5	1.2	35.9	−162.4	30.4	41.2	43.0	31.3	38.4	40.1	37.0	35.7	37.6	35.2	37.0

3

Table 3. Joint coverage at the nominal 95% level for the parameters of the logistic regressions on $\Pr\left(I_{Y_5<0.1}\right)$ ($n = 100$) for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LMlog	PMM	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1	95.4	97.2	96.0	97.4	97.1	97.1	97.5
Model 2	93.7	95.4	94.5	94.3	95.0	94.5	95.3
Model 3	94.9	91.9	92.8	94.3	93.1	93.3	94.9

Table 4. The first two columns show the pooled estimated coefficients and standard errors (SE) from the linear regressions on Y_2 for Models 1 and 2 or Y_4 for Model 3 ($n = 100$) using the full datasets (FD). The remaining columns show the average relative differences (%) as compared to those from FD for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	FD		CC		LMlog		PMM		QRTs (1)		QRTa (1)		QRTs (2)		QRTa (2)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Model 1																
θ_0	0.2	0.2	9.6	23.8	−4.6	7.9	−5.9	4.1	0.5	3.8	1.9	4.0	−9.1	6.4	−0.9	4.7
θ_1	1.9	0.3	−1.5	6.1	−8.3	27.8	−0.9	4.3	0.4	4.3	−0.7	5.4	−3.4	8.4	−1.6	4.8
θ_2	−0.6	0.1	−0.6	14.3	−1.5	10.9	−0.6	3.7	0.1	3.4	0.2	3.9	−1.6	6.2	−0.4	4.3
θ_3	0.3	0.1	−1.6	10.2	3.9	8.4	0.5	2.8	−0.3	2.6	0.7	3.0	0.6	4.7	0.5	3.3
Model 2																
θ_0	$2.0 \cdot 10^{-3}$	0.2	−384.6	24.7	−1954.9	3.7	−660.1	1.4	−451.2	1.0	−555.3	1.1	−431.0	0.9	−484.4	1.0
θ_1	0.5	0.2	−10.0	5.1	−46.0	13.0	−7.7	5.4	−2.7	5.8	−3.8	6.1	−2.3	5.8	−3.9	5.7
θ_2	−0.5	0.1	−1.7	15.2	−5.6	3.7	−1.1	1.2	−0.7	0.8	−0.8	1.0	−0.7	0.8	−0.8	0.9
θ_3	0.5	0.1	0.2	10.4	3.0	3.3	0.7	1.1	0.2	1.0	0.3	1.0	0.1	1.0	0.4	0.9
Model 3																
θ_0	1.1	0.1	16.7	−4.9	−3.0	−9.1	−0.4	28.0	0.5	29.9	1.0	30.1	0.5	30.1	1.4	32.0
θ_1	−0.1	0.2	−29.8	−3.4	−68.5	−20.3	−15.6	38.2	−2.3	39.5	5.2	38.8	−2.8	39.6	10.8	40.9

Table 5. Joint coverage at the nominal 95% level for the parameters of the linear regressions on Y_6 ($n = 100$) for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LMlog	PMM	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1	94.8	95.9	95.0	95.2	95.5	95.8	95.8
Model 2	93.6	90.2	94.2	93.7	93.8	94.3	94.0
Model 3	72.6	95.4	92.9	92.9	93.8	92.7	93.5

Table 6. Average relative bias (%) at different probabilities of the empirical cumulative distribution of Y_5 ($n = 1000$) for the complete case analysis (CC), normal imputation (LM), log-normal imputation (LMlog), predictive mean matching (PMM), linear quantile regression imputation (QR), and imputation based on the symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LM	LMlog	PMM	QR	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1									
0.01	−32.3	431.1	−37.7	−7.3	344.7	−1.7	−12.8	17.6	3.2
0.05	−25.2	59.8	−19.7	−0.9	45.6	0.1	−7.0	5.7	−1.9
0.25	−16.0	−4.6	0.8	−0.1	−4.6	0.1	−1.4	0.9	−0.8
0.5	−9.6	−4.5	2.1	−0.1	−3.3	0.0	0.3	0.0	0.3
0.75	−4.5	−0.2	0.5	−0.0	−0.1	0.0	0.3	−0.1	0.3
0.95	−0.9	0.1	−0.3	−0.0	0.1	0.0	0.0	−0.1	−0.0
Model 2									
0.01	14.2	461.5	0.5	16.1	141.1	2.1	1.7	1.9	2.2
0.05	5.3	82.9	3.8	12.4	26.4	1.1	0.5	1.2	0.8
0.25	−5.0	0.1	1.9	4.1	−0.7	0.4	−1.2	0.4	−1.3
0.5	−6.4	−6.5	−0.1	0.3	−3.0	0.1	−0.9	0.1	−1.0
0.75	−4.3	−0.5	−1.3	−0.6	−0.4	−0.0	0.0	−0.0	0.0
0.95	−1.0	−0.1	−1.8	−0.2	0.0	−0.0	0.0	−0.0	0.0
Model 3									
0.01	2.9	119.1	−19.8	1.0	6.7	2.9	2.6	3.5	5.5
0.05	3.0	9.9	−11.2	1.7	1.7	0.9	0.6	1.6	1.9
0.25	2.9	−7.8	12.1	0.3	0.9	0.5	0.5	0.6	0.6
0.5	2.5	0.3	7.5	0.6	0.5	0.4	0.3	0.4	0.3
0.75	1.7	3.4	0.8	0.4	0.1	0.3	0.1	0.3	0.1
0.95	0.7	−0.5	−4.0	0.2	−0.1	0.2	0.2	0.2	0.1

Table 7. The first two columns show the pooled estimated coefficients and standard errors (SE) from the logistic regressions on $\Pr\left(I_{Y_5<0.1}\right)$ ($n = 1000$) using the full datasets (FD). The remaining columns show the average relative differences (%) as compared to those from FD for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	FD		CC		LMlog		PMM		QRTs (1)		QRTa (1)		QRTs (2)		QRTa (2)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Model 1																
θ_0	3.2	0.3	0.4	23.5	3.0	22.6	−0.3	21.9	0.2	22.0	−5.5	22.6	2.0	20.5	−4.6	22.0
θ_1	−1.8	0.2	0.6	15.7	1.1	15.9	−0.0	15.3	0.3	15.4	−4.0	14.1	1.4	14.4	−3.7	13.8
θ_2	−1.8	0.2	0.7	13.9	−2.7	12.8	0.0	13.4	0.1	13.8	−4.6	12.7	0.5	14.0	−4.8	12.2
θ_3	−0.2	0.1	−1.4	13.9	−3.7	14.4	−1.7	13.8	−1.3	13.7	−6.8	12.9	−1.4	13.8	−5.1	12.7
Model 2																
θ_0	1.1	0.2	−12.9	23.0	−5.7	22.0	9.8	27.0	1.0	21.6	−12.9	21.2	1.2	21.5	−13.1	22.3
θ_1	−0.5	0.1	−8.4	13.3	−4.8	12.8	6.3	15.6	0.6	13.5	−8.0	13.3	0.7	14.2	−8.1	13.5
θ_2	−0.7	0.1	−8.0	8.9	−10.7	9.8	−6.8	10.5	0.1	10.2	−5.2	10.0	−0.0	10.7	−4.9	10.5
θ_3	−0.1	0.1	−8.1	9.7	−9.6	10.8	−7.5	11.2	−0.3	10.9	−6.2	10.7	0.5	11.5	−6.0	10.6
Model 3																
θ_0	−1.5	0.2	−0.8	48.7	−27.0	36.2	2.2	147.8	−0.8	53.9	−0.7	57.9	−1.0	56.7	−1.1	57.5
θ_1	0.2	0.1	−5.7	34.1	−158.8	27.2	10.6	111.7	−5.9	38.4	−5.3	41.2	−6.9	40.3	−7.4	40.9

∞

Table 8. Joint coverage at the nominal 95% level for the parameters of the logistic regressions on $\Pr\left(I_{Y_5<0.1}\right)$ ($n = 1000$) for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LMlog	PMM	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1	94.7	97.0	95.9	97.3	96.0	96.5	96.1
Model 2	93.5	95.3	95.0	96.2	94.2	95.7	95.2
Model 3	94.2	57.2	94.3	93.2	93.2	93.4	93.3

Table 9. The first two columns show the pooled estimated coefficients and standard errors (SE) from the linear regressions on Y_2 for Models 1 and 2 or Y_4 for Model 3 ($n = 1000$) using the full datasets (FD). The remaining columns show the average relative differences (%) as compared to those from FD for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	FD		CC		LMlog		PMM		QRTs (1)		QRTa (1)		QRTs (2)		QRTa (2)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Model 1																
θ_0	0.2	0.1	11.4	23.8	−6.1	13.1	−0.5	4.1	1.1	3.6	0.9	4.1	−0.0	4.2	1.1	4.3
θ_1	1.9	0.1	−1.3	5.8	−12.0	110.5	−0.1	3.8	0.2	4.2	−0.8	4.9	−0.2	4.3	−1.0	5.2
θ_2	−0.6	$2.7 \cdot 10^{-2}$	−0.4	14.2	−2.2	25.1	−0.0	3.7	0.1	3.5	0.3	3.8	−0.1	4.0	0.2	4.1
θ_3	0.3	$1.6 \cdot 10^{-2}$	−1.7	9.8	5.5	36.3	0.1	2.5	−0.1	2.5	1.1	2.8	−0.1	2.8	1.0	3.0
Model 2																
θ_0	$2.3 \cdot 10^{-2}$	0.1	−33.8	24.9	−164.9	3.6	−22.9	1.2	−0.6	0.6	−25.0	0.8	−0.9	0.6	−26.7	0.8
θ_1	0.5	0.1	−8.6	5.0	−71.9	52.2	−3.5	5.2	−0.1	5.1	−3.2	5.7	−0.3	5.6	−3.3	6.1
θ_2	−0.5	$3.0 \cdot 10^{-2}$	−1.2	15.3	−7.9	4.0	−0.6	1.0	−0.0	0.6	−0.5	0.8	−0.1	0.7	−0.5	0.8
θ_3	0.5	$1.6 \cdot 10^{-2}$	0.5	10.0	4.7	8.7	0.3	0.9	−0.0	0.8	0.4	0.8	−0.0	0.8	0.4	0.8
Model 3																
θ_0	1.1	$3.2 \cdot 10^{-2}$	16.5	−5.4	−3.8	−22.6	−0.3	89.6	−0.5	30.0	−0.3	30.0	−0.5	28.7	−0.3	29.5
θ_1	−0.1	0.1	−31.6	−3.9	−70.0	−49.0	−9.7	113.8	−8.7	40.0	−6.0	39.7	−8.2	38.2	−5.4	39.0

Table 10. Joint coverage at the nominal 95% level for the parameters of the linear regressions on Y_6 ($n = 1000$) for the complete case analysis (CC), log-normal imputation (LMlog), predictive mean matching (PMM), and symmetric (QRTs) and asymmetric (QRTa) transformed quantile regression imputation models. The latter were fitted with either (1) known or (2) unknown λ_p.

	CC	LMlog	PMM	QRTs (1)	QRTa (1)	QRTs (2)	QRTa (2)
Model 1	94.5	93.0	94.7	94.8	94.6	95.0	94.8
Model 2	94.5	62.1	94.5	95.0	94.8	95.2	94.5
Model 3	42.0	34.2	90.3	93.5	94.0	93.2	93.4
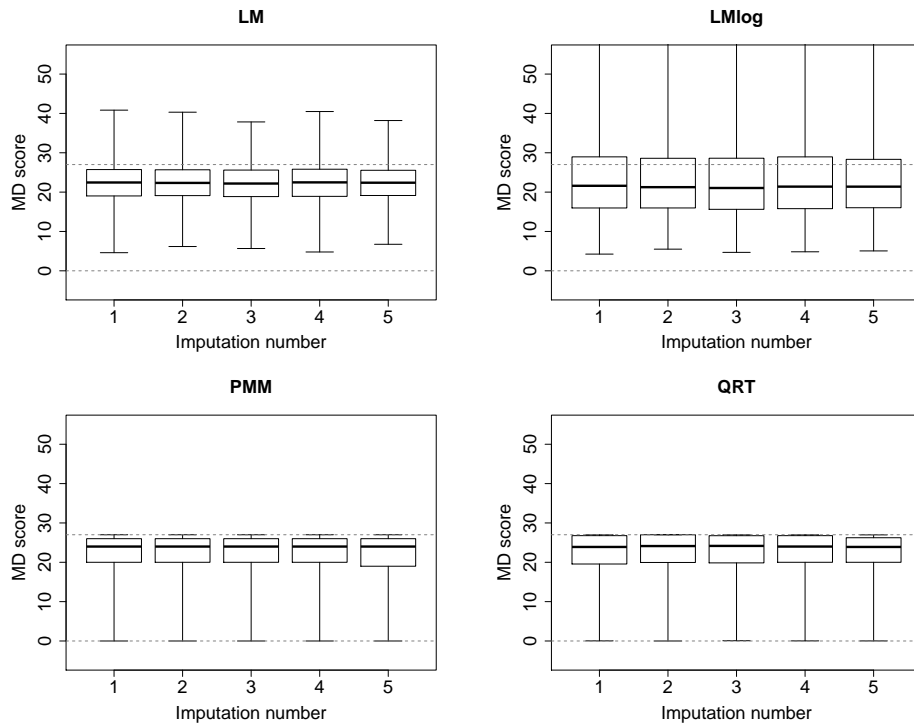
Figure 1. Boxplots of imputed mathematical development (MD) scores (Celtic Country Teacher Survey) by imputation number, with whiskers extending to the minimum and maximum values, for normal imputation (LM), log-normal imputation (LMlog), predictive mean matching (PMM), and imputation based on the transformed quantile regression model (QRT). The horizontal dashed lines mark the theoretical bounds of the scores.
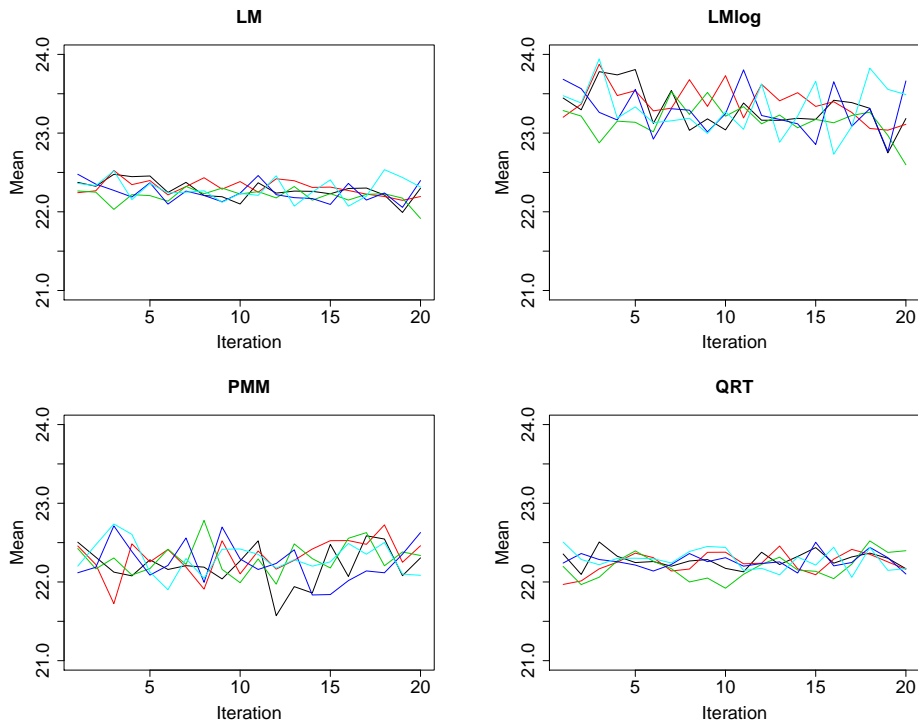
Figure 2. Stream plots of the Monte Carlo chains (mean) resulting from the normal imputation (LM), log-normal imputation (LMlog), predictive mean matching (PMM), and imputation based on the transformed quantile regression model (QRT) of the mathematical development scores (Celtic Country Teacher Survey) for 20 Gibbs sampler's iterations and 5 imputations (streams).
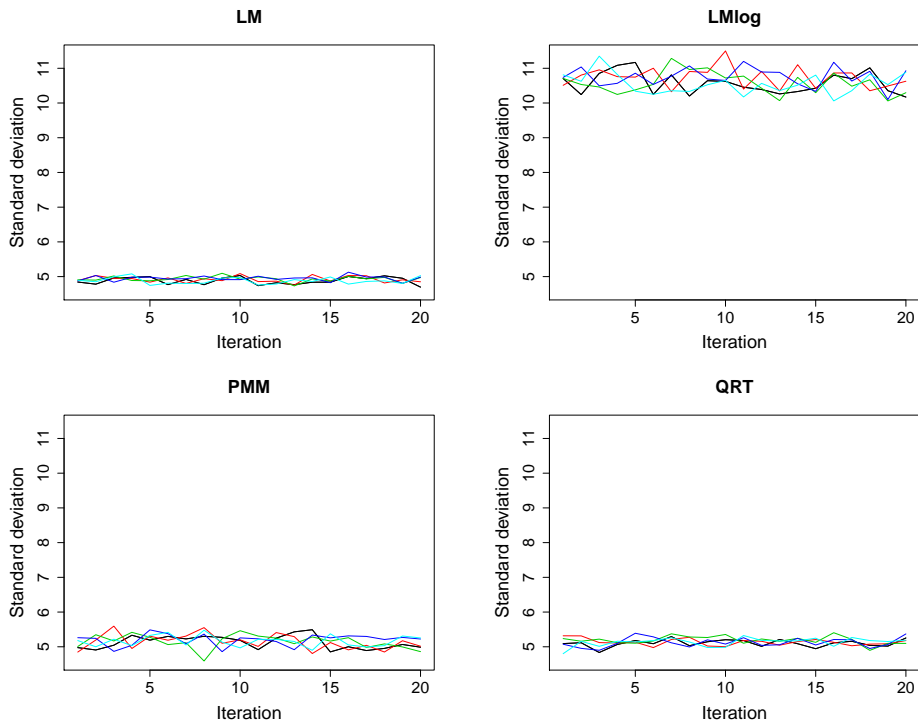
Figure 3. Stream plots of the Monte Carlo chains (standard deviation) resulting from the normal imputation (LM), log-normal imputation (LMlog), predictive mean matching (PMM), and imputation based on the transformed quantile regression model (QRT) of the mathematical development scores (Celtic Country Teacher Survey) for 20 Gibbs sampler's iterations and 5 imputations (streams).