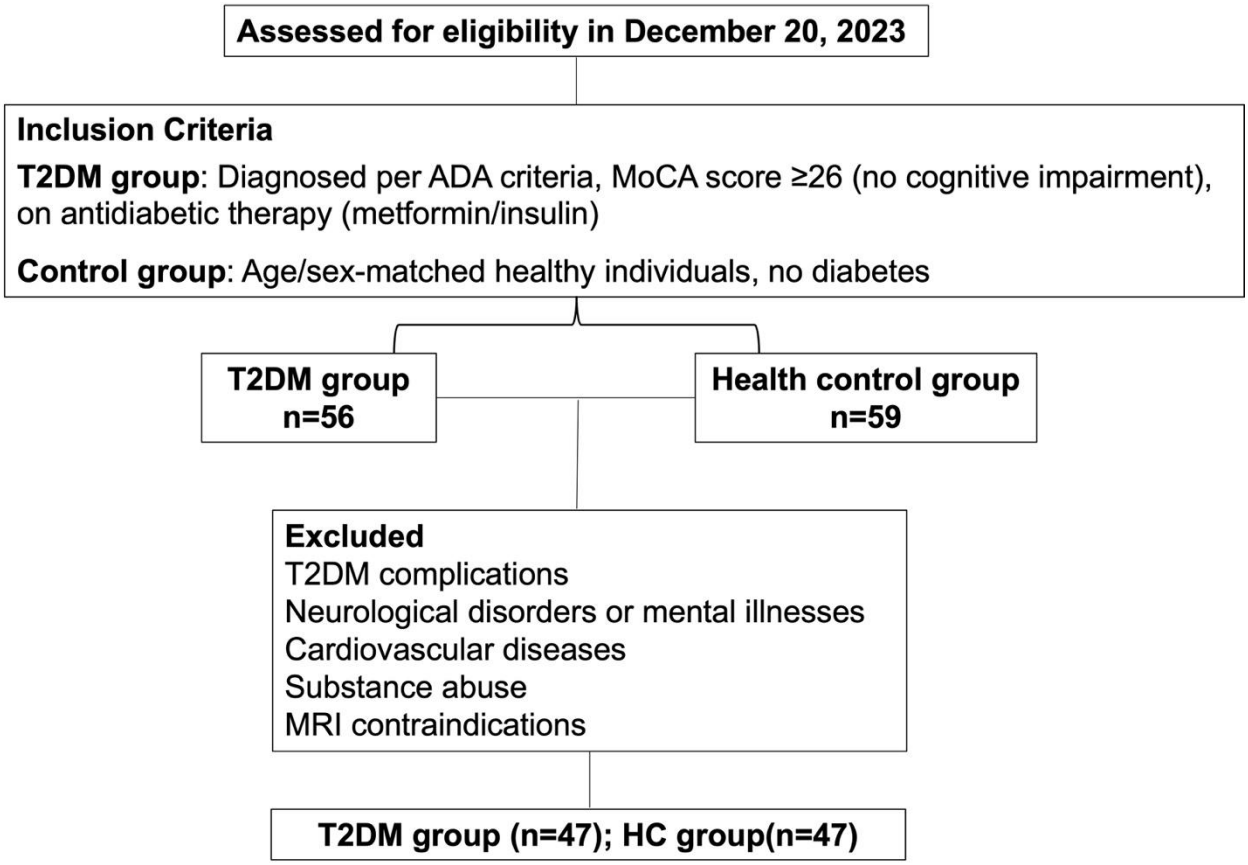
Supplementary material

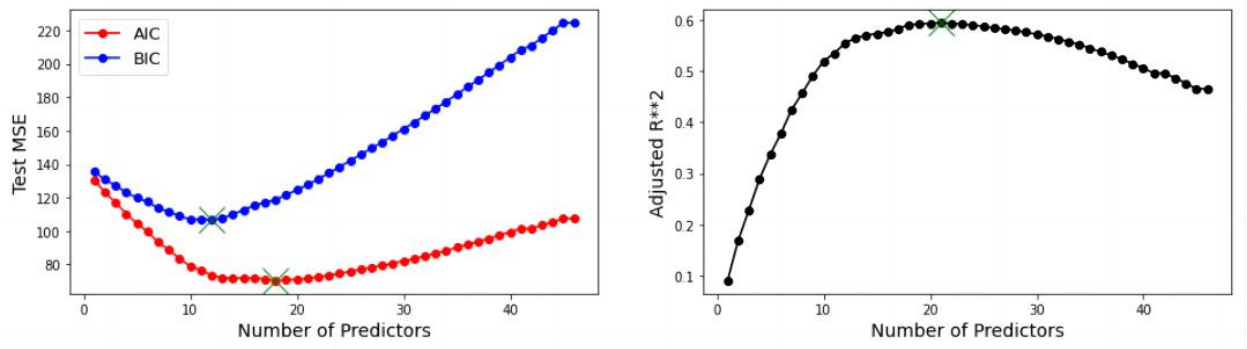
Model evaluations

To evaluate the classification performance, we applied the stratified 10-fold cross-validation strategy (Pereira et al., 2009). This means that the model was trained on 9 folds of data at a time, while the remaining one-fold was kept for validation. The training and evaluation process was repeated 10 times. During each iteration of the cross-validation, evaluation metrics including accuracy, precision, recall, and F1-score were computed. Since each training and evaluation iteration can yield a set of metrics, we report the average and variance of these metrics to provide a comprehensive assessment of the model's performance.

To evaluate the disease duration prediction performance, we employed the Leave-One-Out Cross-Validation (LOOCV) method (Pereira et al., 2009), due to the limited sample size in the patient group. This method is suitable for evaluating prediction performance on relatively small sample sizes and involves iteratively training the model on all samples except one and evaluating its performance on the left-out sample. Then, the Root Mean Square Error (RMSE) and its variance in the LOOCV are used to evaluate the performance of each model.

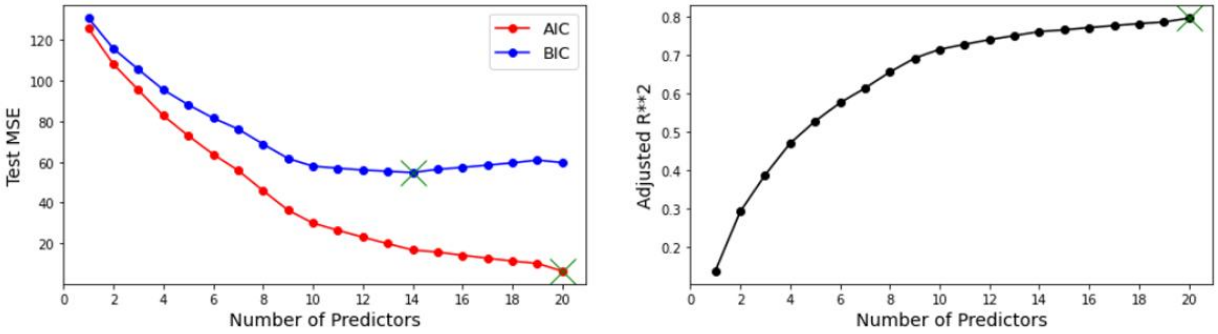


Supplementary Figure 1 STROBE diagram of participant flow. T2DM: Type 2 diabetes mellitus.

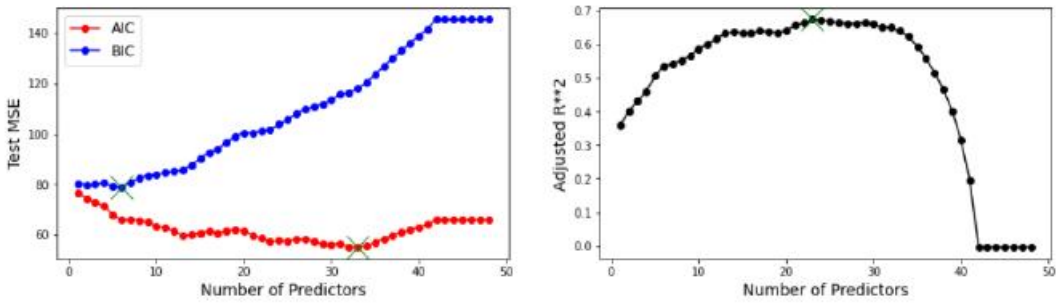


Supplementary Figure 2 AIC, BIC and Adjusted R-squared during the forward feature selection for advanced connectivity model on T2DM classification. T2DM, Type 2 diabetes mellitus; AIC (Akaike Information Criterion) = -2 * log-likelihood + 2 *

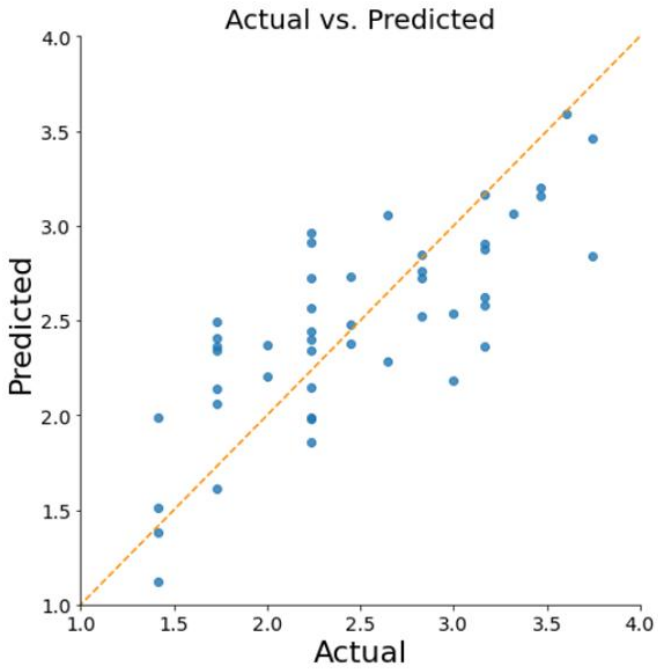
k, BIC (Bayesian Information Criterion) = -2 * log-likelihood + k * log(n). Both metrics are model selection criteria that integrate model performance and penalize large models. The smaller the values of these two metrics, the better the overall classification performance and generalization of the model. Adjusted R-squared = 1 - (1 - R-squared) * (n - 1) / (n - k - 1)) which is a statistical measure that assesses the goodness of fit of a model. It is an adjusted version of the R-squared (coefficient of determination) that considers the number of predictors in the model. Based on the three estimates of test error, it is suggested that a model with 10-20 predictors is needed. Subsequently, models were tested using these subsets, and it was found that the model with 18 selected features demonstrated the best performance in terms of accuracy and precision during cross-validation. The detailed report on the selected 18 features and the performance of the optimal model can be found in the main text.



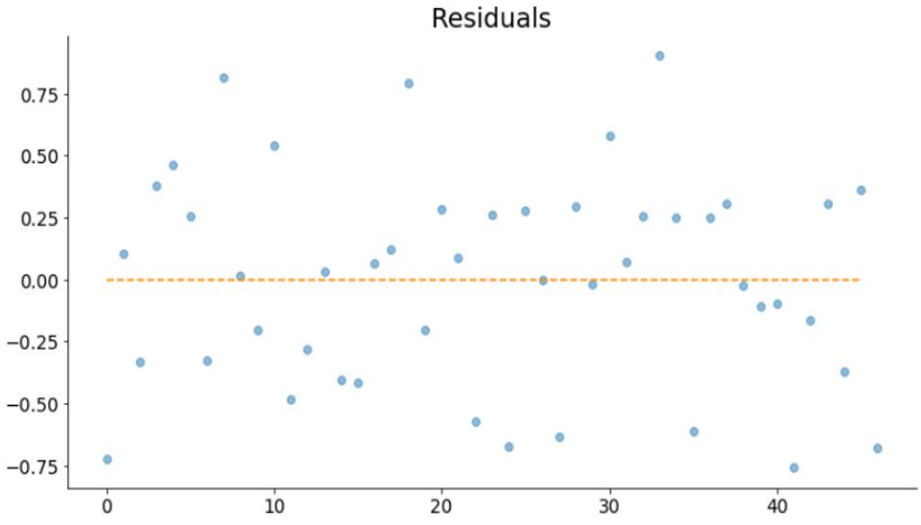
Supplementary Figure 3 AIC, BIC and Adjusted R-squared during the forward feature selection of the connectivity interaction terms for T2DM classification. T2DM, Type 2 diabetes mellitus; AIC (Akaike Information Criterion) = -2 * log-likelihood + 2 * k, BIC (Bayesian Information Criterion) = -2 * log-likelihood + k * log(n), Adjusted R-squared = 1 - (1 - R-squared) * (n - 1) / (n - k - 1)). See detailed introduction for the model evaluation metrics in Supplementary Figure 1. We consider the second-order interaction, and the problem here is that the 18 selected features can yield 153 second-order interaction terms. Therefore, we need to select the terms that contribute significantly to the classification from these 153 interaction terms. A logistic regression function is then fitted with the interaction terms, and the forward stepwise selection method is applied. This figure reports the AIC, BIC and adjusted R-squared for the best subset with different numbers of predictors (we restricted the maximum number of terms to 20). These metrics suggest that the models with 12-20 interaction terms can fit the data well. By using the original 18 connection features with different subsets of interaction terms (12-20) to build the model, we find that the model including 14 selected interaction terms gives the best classification performance (the results of optimal model are reported in main text).



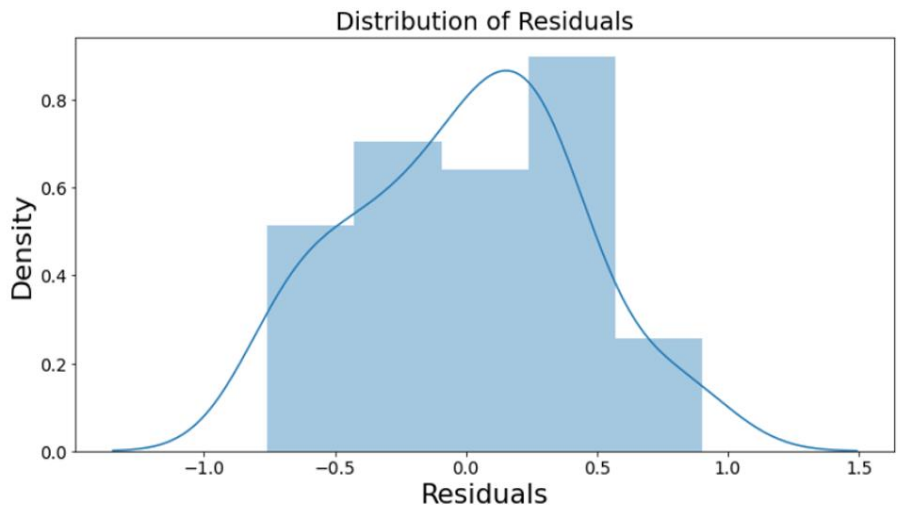
Supplementary Figure 4 AIC, BIC and Adjusted R2 during the forward feature selection of the interaction terms for T2DM disease duration prediction. T2DM, Type 2 diabetes mellitus; AIC (Akaike Information Criterion) = -2 * log-likelihood + 2 * k, BIC (Bayesian Information Criterion) = -2 * log-likelihood + k * log(n), Adjusted R-squared = 1 - (1 - R-squared) * (n - 1) / (n - k - 1)). See detailed introduction for the model evaluation metrics in Supplementary Figure 1.



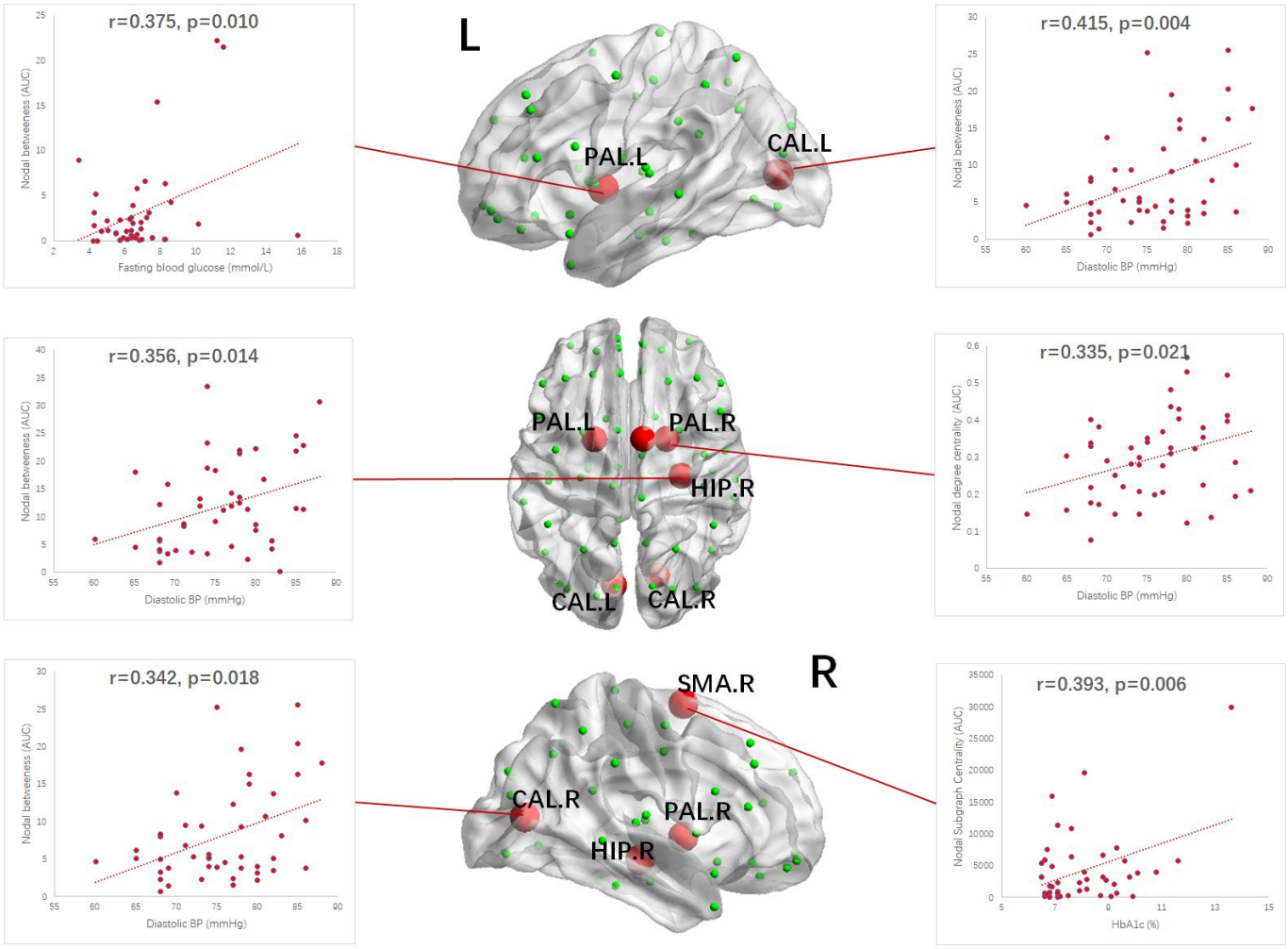
Supplementary Figure 5 Graph of predicted values versus the ground truth values of the regression model for T2DM (Type 2 diabetes mellitus) disease duration.



Supplementary Figure 6 Scatter plot of the residuals of the regression model for T2DM (Type 2 diabetes mellitus) disease duration.



Supplementary Figure 7 Distribution of the residuals of the regression model for T2DM (Type 2 diabetes mellitus) disease duration.

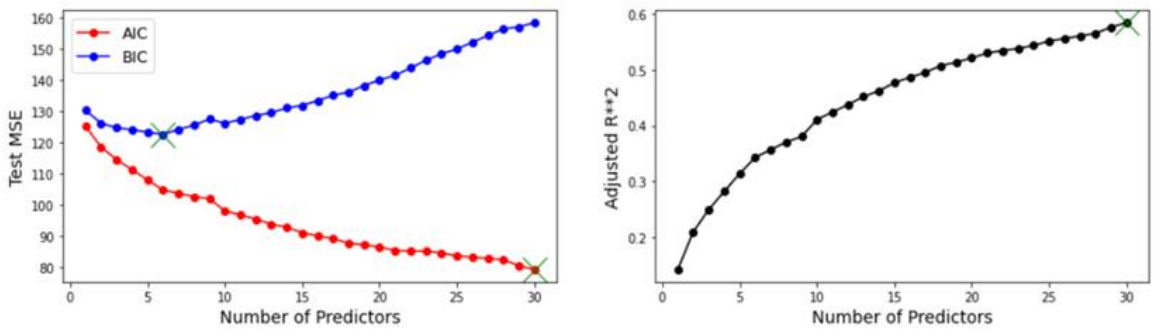


Supplementary Figure 8 Nodal connectivity efficiency positively correlated with T2DM clinical indicators. Each region of interest (ROI) was considered as a network node, and the connectivity between nodes was weighted using the fractional anisotropy (FA) values between ROIs. This allowed us to transform the connectivity information into a weighted brain network consisting of 116 nodes. We then employed graph theory analysis methods[1,2] to calculate nodal connectivity efficiency, including nodal betweenness and nodal degree centrality. Then, we examined the correlation between these connectivity metrics and clinical data. Bilateral pallidum: PAL.R/ PAL.L; Diastolic BP: diastolic blood pressure.

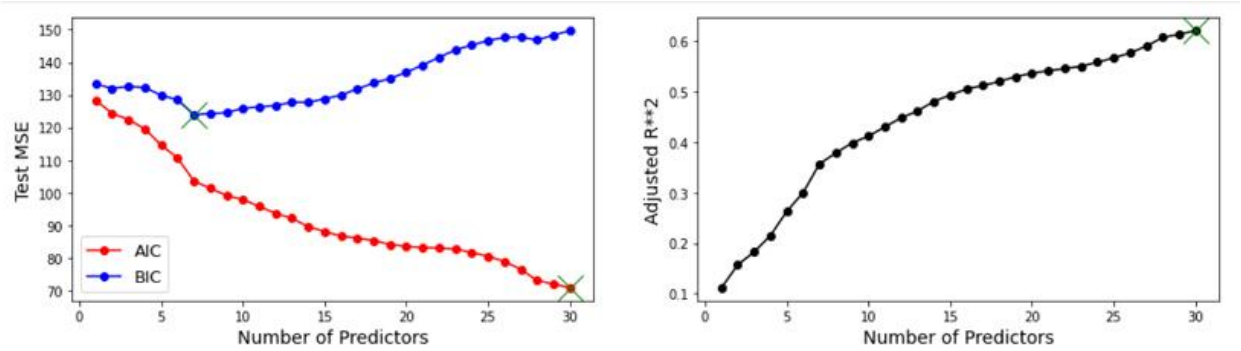
References

1 Yang, H., Chen, X., Chen, Z. B., Li, L., Li, X. Y., Castellanos, F. X., ... & Yan, C. G. (2021). Disrupted intrinsic functional brain topology in patients with major depressive disorder. Molecular psychiatry, 1-9.

2 Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. PLoS computational biology, 3(2), e17.



Supplementary Figure 9 Feature Selection for Classification Models via Surface Size and Voxel Size Surface Size with Feature Selection. AIC, BIC and Adjusted R2 during the forward feature selection for the surface size.



Supplementary Figure 10 Voxel Size with Feature Selection. AIC, BIC and Adjusted R2 during the forward feature selection for the voxel size.

Supplementary Table 1 Performance for the models with different amount of surface size features (values in brackets: variances)

	Accuracy	Precision	Recall	F1-score
N=6	0.7766	0.8109	0.7733	0.7627
	(0.0026)	(0.0100)	(0.0360)	(0.0094)
N=30	0.7860	0.7915	0.8111	0.7914
	(0.0063)	(0.0091)	(0.0131)	(0.0051)
N=90	0.4678	0.4841	0.4733	0.4649
	(0.0057)	(0.0103)	(0.0161)	(0.0056)

Supplementary Table 2 Performance for the models with different amount of voxel size features (values in brackets: variances)

	Accuracy	Precision	Recall	F1-score
N=6	0.7766	0.7694	0.8089	0.7812
	(0.0015)	(0.0045)	(0.0106)	(0.0019)
N=30	0.6801	0.6950	0.6978	0.6882
	(0.0127)	(0.0262)	(0.0086)	(0.0111)
N=90	0.3614	0.3478	0.3844	0.3610
	(0.0057)	(0.0111)	(0.0274)	(0.0179)