SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Predicting Compositional Time Series via Autoregressive Dirichlet Estimation

Ganbin ZHOU[1,2], Ping LUO[1,2] & Qing HE[1,2*]

[1] *Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing* 100190*, China;*
[2] *University of Chinese Academy of Sciences, Beijing* 100049*, China.*

Appendix A Experimental Details

Appendix A.1 Data Preparation

We prepare two data sets of real-world applications to evaluate our method, one is about CCEE, the other is about the World Development Indicators (WDI), e.g., GDP, Population and Export.

Table A1 Data set of CCEE (math).

No.	Region	Type	Start	End		
1	Beijing	Art	2002	2015	1.Algebra	2.Probability theory
2	Shandong	Art	2005	2015	3.Trigonometry	4.Planimetry
3	Shanghai	Art	2000	2015	5.Stereometry	6.Logic & Inference
					7.Algorithms	8.Calculus

Chinese College Entrance Examination. The first data set is CCEE data set. We collect CCEE papers[1)] of mathematics exams from Beijing, Shandong and Shanghai. The detailed information of this data set is shown in the Table A1. The components of each vector represent the proportions of knowledge points in a specific type of exam paper over time. Note that papers of Beijing doesn't contain calculus, while those of Shandong and Shanghai do.

World Development Indicators. The second data set is collected from the World Bank[2)], which contains seven indicators. Each indicator is collected from different countries over time, and the description is detailed in Table A2.

Table A2 Data sets of the World Development Indicators.

No.	Indicator	Year		Country Codes
		Start	End	
1	GDP	1968	2014	ARB, CHN, EUU, JPN, USA, Others
2	Population	1960	2014	ARB, CHN, EUU, JPN, USA, Others
3	Export	1960	2014	ARB, CHN, EUU, JPN, USA, Others
4	CO_2	1960	2011	ARB, CHN, EUU, IND, JPN, USA, Others
5	Energy	1971	2011	ARB, CHN, EUU, IND, JPN, USA, Others
6	Reserves	1977	2014	BRA, CHE, CHN, JPN, KOR, SAU, USA

* "Others" in the Country Codes column means the rest of the world.

For each year, there is only one vector for an indicator, each of its components represents a proportion, relative to the total of the regions. For example, there is a vector in the GDP indicator data set every year, each component of this vector represents the GDP proportion corresponding to a region, relative to the total GDP of ARB, CHN, EUU, JPN, USA and the others.

* Corresponding author (email: heqing@ict.ac.cn)
1) http://www.jyeoo.com/
2) http://data.worldbank.org/indicator

Appendix A.2 Evaluation Methods and Implementation Details

We compare our method ADE with three state-of-the-art baselines, two special designed methods:

- The first is the DRHT [3], which uses a 3-order polynomial regression to model the data after the hyperspherical transformation.
- The second is CDES [2], in which the smooth strength is set as 0.8 in the experiments.
- The third is VARMA [1]. Since the data set used in this paper is high-dimensional and insufficient training data for VARMA, we choose low-order VARMA model.
- There are two special designed methods OA and OD. OA only maximizes the Autoregression Objective. OD only maximizes the Dirichlet Objective.

We bisect each data set. The first part is used for training, and the second is used for testing. For a sequence $\{\boldsymbol{\theta}_i\}_{i=1}^{t-1}$, assuming the predictions from one method are $\{\hat{\boldsymbol{\theta}}_i\}_{i=\lceil t/2\rceil}^{t-1}$, then the prediction error is defined as $\|\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i\|_1$. In each sequence, we compare the average prediction errors from different methods. For all data sets, we set $l = 6$, $s = 4$. For WDI data set, we set $m = 10$ after some preliminary experiments. While for CCEE data set, we set $m = 4$ due to its shortage (there are only 10 or 15 vectors for the sequences).

Appendix A.3 Experimental Results

The experimental results of *Chinese College Entrance Examination* are shown in Table A3.

Table A3 The average errors of different methods on CCEE.

Dataset	OD	ADE			OA	DRHT	CDES	VARMA
		$\sigma=10^1$	$\sigma=10^{-1.5}$	$\sigma=10^{-2}$				
1	0.1637	0.1634	0.1649	**0.1582**	0.1611	0.5585	0.1746	0.6164
2	0.1382	0.1381	0.1406	0.1440	0.1480	0.5581	**0.1288**	1.5762
3	**0.2074**	0.2079	0.2437	0.2638	0.2676	0.7068	0.2522	0.6441

Observed from the results in Table A3, ADE outperforms DRHT, CDES and VARMA, except that CDES is slightly better than ADE on data set 2. Furthermore, based on two-tailed Student's t-test, there are statistically significant differences between the errors of ADE and those of DRHT, CDES and VARMA on data set 1 and 3. Then, based on Table A3, we compare OA and OD with ADE. The results are shown in Figure A1.

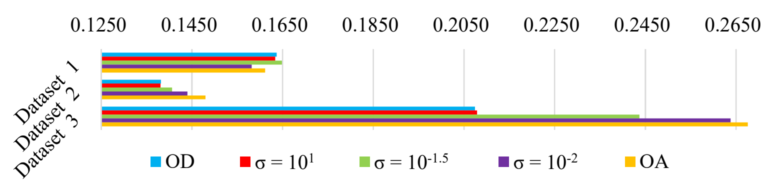

Figure A1 Errors of OA, OD and ADE with different σ on CCEE.

Observed from Figure A1, OD outperforms OA in data set 2 and 3. When σ becomes bigger ($\frac{1}{2\sigma^2}$ becomes smaller), the performance of ADE is closer to OD, and vice versa. Every year the Chinese government promulgates an official outline which prescribes a general proportion of CCEE knowledge points. Empirically, there are tiny changes in this outline every year. We believe that the knowledge proportions of current year CCEE depends more on this outline than previous papers. It explains why OD performs better than OA in most CCEE data sets. However, since there are some CTS in CCEE where OA performs better, it is difficult to determine whether the CCEE paper next year will be independent of previous papers or follow some regularity. In general, $\sigma = 10^1$ will be a good choose for CCEE prediction.

The experimental results of *World Development Indicators* are shown in Table A4.

Table A4 The average errors $(\times 10^{-3})$ of different methods on WDI.

Dataset	OD	ADE			OA	DRHT	CDES	VARMA
		$\sigma=10^{-2}$	$\sigma=10^{-2.5}$	$\sigma=10^{-4}$				
GDP	64.20	52.50	43.42	**41.77**	**41.77**	102.48	45.34	276.41
Population	7.34	7.03	5.86	**0.35**	**0.35**	0.57	3.91	17.50
Export	22.14	23.70	20.29	18.05	18.05	45.76	**17.90**	428.67
CO2	39.75	35.64	27.48	**24.98**	25.00	49.26	26.39	343.26
Energy	31.18	28.07	19.70	**15.49**	15.50	25.42	19.34	407.86
Reserves	156.22	90.29	85.00	**85.29**	**85.29**	218.72	95.55	1072.78

Observed from the results in Table A4, VARMA achieves the worst performance. The reason is that the number of parameters of the VARMA model is more than the square of the number of dimensions (bigger than 36), so the number of training data is not enough for VARMA model, while ADE, DRHT and CDES can handle the data, even the dimension is high. Meanwhile, ADE outperforms DRHT, CDES and VARMA, except that CDES is slightly better than ADE on Export. Furthermore, based on two-tailed Student's t-test, there are statistically significant differences between the errors of ADE and those of the three baselines, except Export. Again, based on Table A4, we compare OA and OD with ADE. The results are shown in Figure A2.

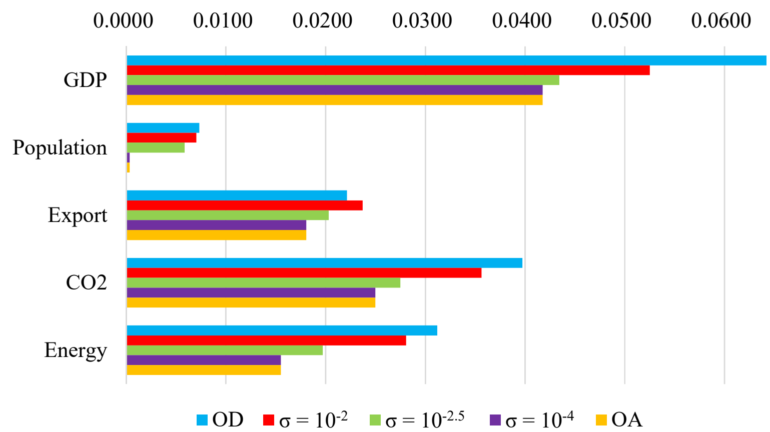

Figure A2 Errors of OA, OD and ADE with different σ on WDI.

Observed from Figure A2, when σ becomes smaller ($\frac{1}{2\sigma^2}$ becomes bigger), the performance of ADE is closer to OA. Since these financial CTS in experiments change regularly, the experiments indicate that for such kind of CTS which changes regularly, we should set σ to be smaller.

In summary, all the results of these data sets validate the effectiveness and robustness of the ADE method.

References

1 Mills T. Forecasting compositional time series. Quality & Quantity, 2010, 44: 673-690.

2 Raharjo H, Xie M, Brombacher A. On modeling dynamic priorities in the analytic hierarchy process using compositional data analysis. European Journal of Operational Research, 2009, 194: 834-846.

3 Wang H, Liu Q, Mok H, et al. A hyperspherical transformation forecasting model for compositional data. European journal of operational research, 2007, 179:459-468.