

2022.06. 대한산업공학회 추계학술대회

딥 클러스터링을 통한 가계 금융 다양성 분석

황윤태 (UNIST 산업공학과)

이용재 (UNIST 산업공학과)

Frank J. Fabozzi (EDHEC Business School)

이 용 재

금융공학연구실

산업공학과

UNIST
ULSAN NATIONAL INSTITUTE OF
SCIENCE AND TECHNOLOGY

연구 배경

■ 금융경제학 연구에서의 가계금융 다양성

- Consumption-based CAPM: 소비와 portfolio allocation을 연결지음
 - Merton (1973), Lucas (1978), Breeden (1979, 1986), Grossman and Shiller (1982) 등
- Equity premium puzzle (Mehra and Prescott, 1985): Representative consumer를 기반으로 한 consumer-based CAPM으로는 equity premium이 작게 측정된다
- 이를 위해 income, stock holding 여부, savings, wealth등의 다양성을 모형에 반영하기 시작
 - Mankiw and Zeldes (1991), Lucas (1994), Heaton and Lucas (1997), Krueger, Mitman, and Perri (2016) 등

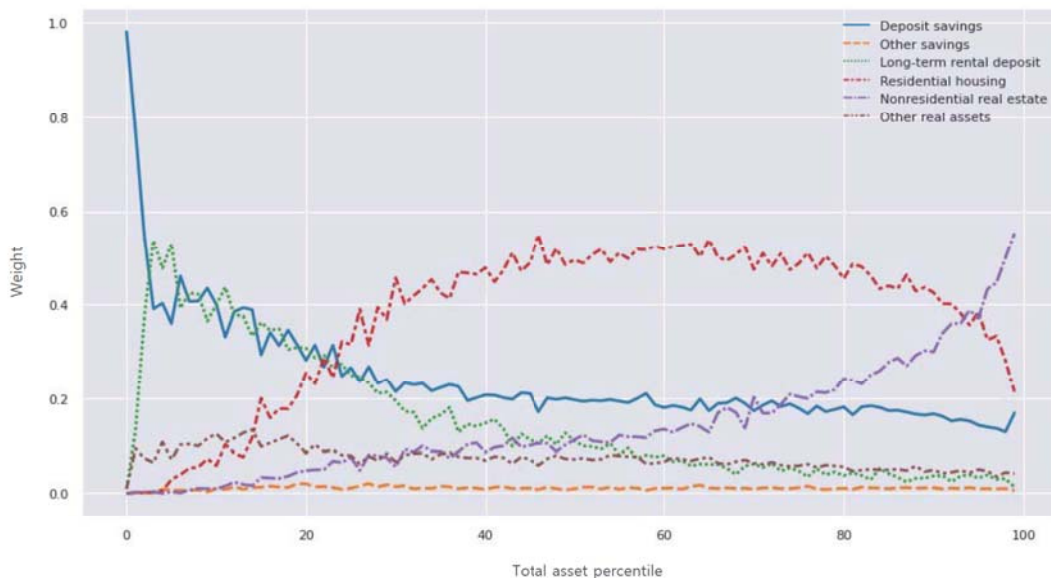
연구 배경

- Krueger, Mitman, and Perri (2016):

“실제 데이터에서 우리가 관측하는
wealth, income, expenditure의 joint distribution을
더 정확하게 모델에 반영하기 위해서는
더 많은 차원의 heterogeneity를 고려해야한다”

연구 배경

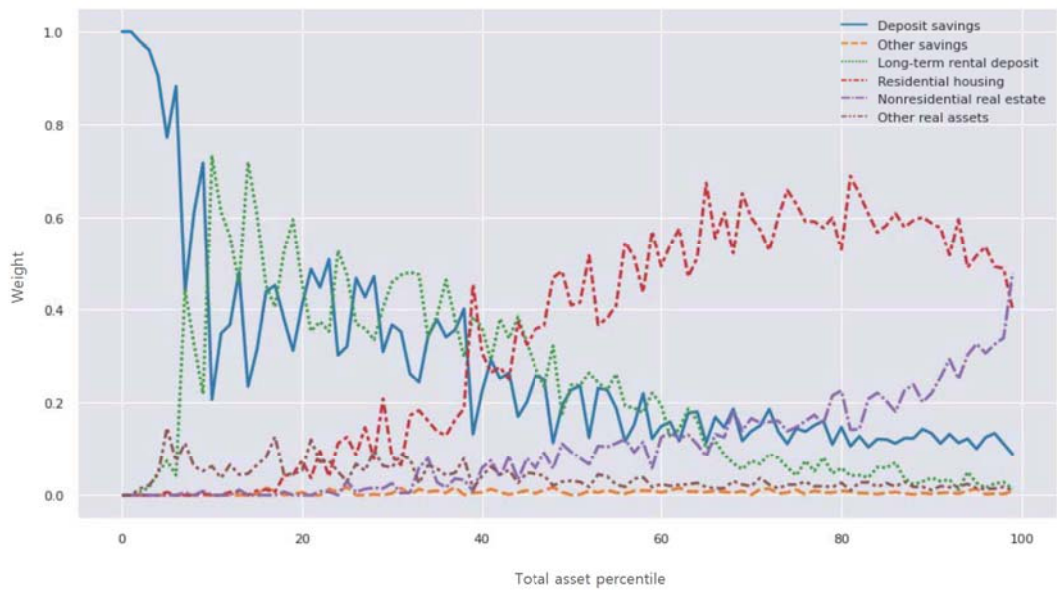
2017~2020 가계 재무 포트폴리오



자산 총액이 증가함에 따라 유형별 자산의 비중은 비선형적으로 변화

연구 배경

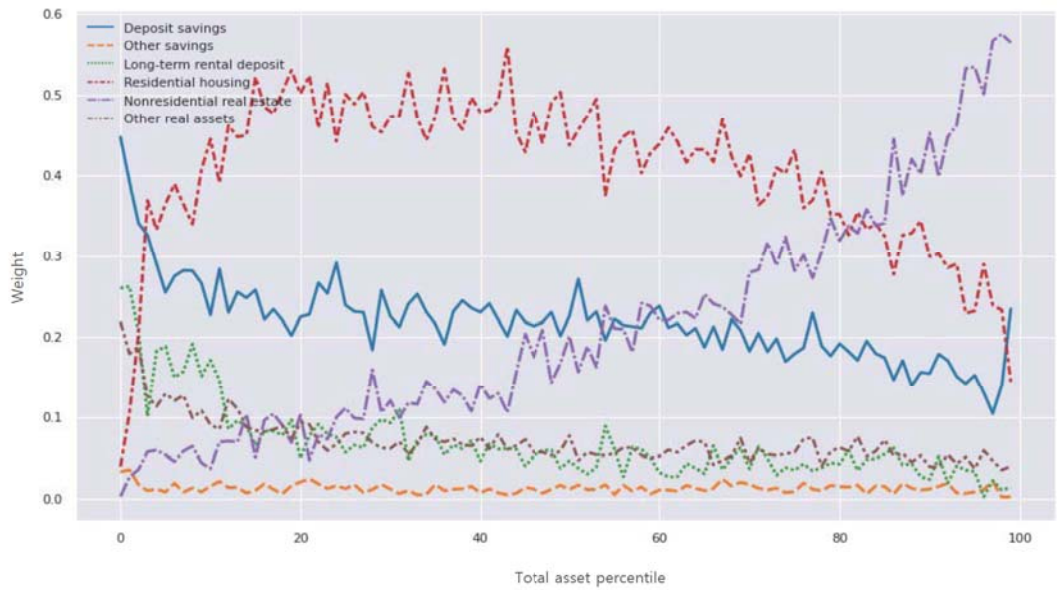
2017~2020 소득 1분위 가계 재무 포트폴리오



자산 총액이 증가함에 따라 유형별 자산의 비중은 비선형적으로 변화

연구 배경

2017~2020 소득 5분위 가계 재무 포트폴리오

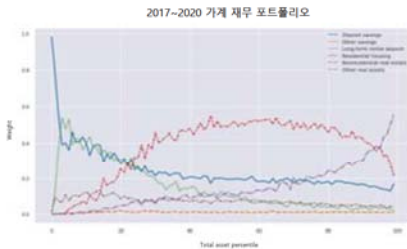


자산 총액이 증가함에 따라 유형별 자산의 비중은 비선형적으로 변화

연구 배경

과거의 접근법

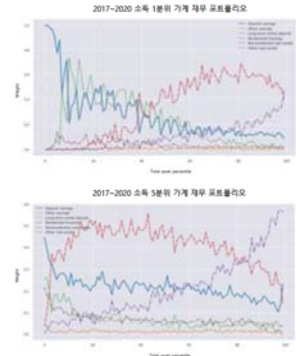
가계금융 데이터



소득에 따른 분류



결과 분석



- 한두개의 항목 별로 분류하고 분석
- 분류를 하고 난 뒤에도 그룹 내의 heterogeneity가 많이 사라지지 않음

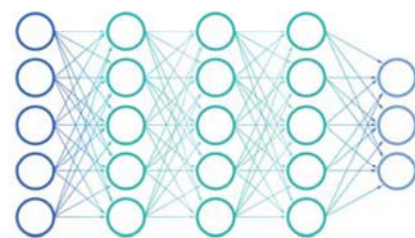
연구 배경

본 연구의 접근법

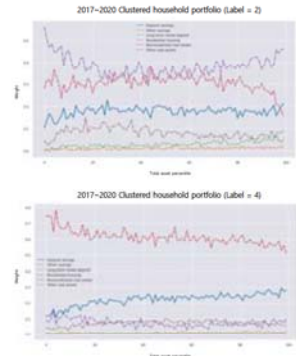
가계금융 데이터



모든 데이터를 기반으로
딥러닝을 통한 분류



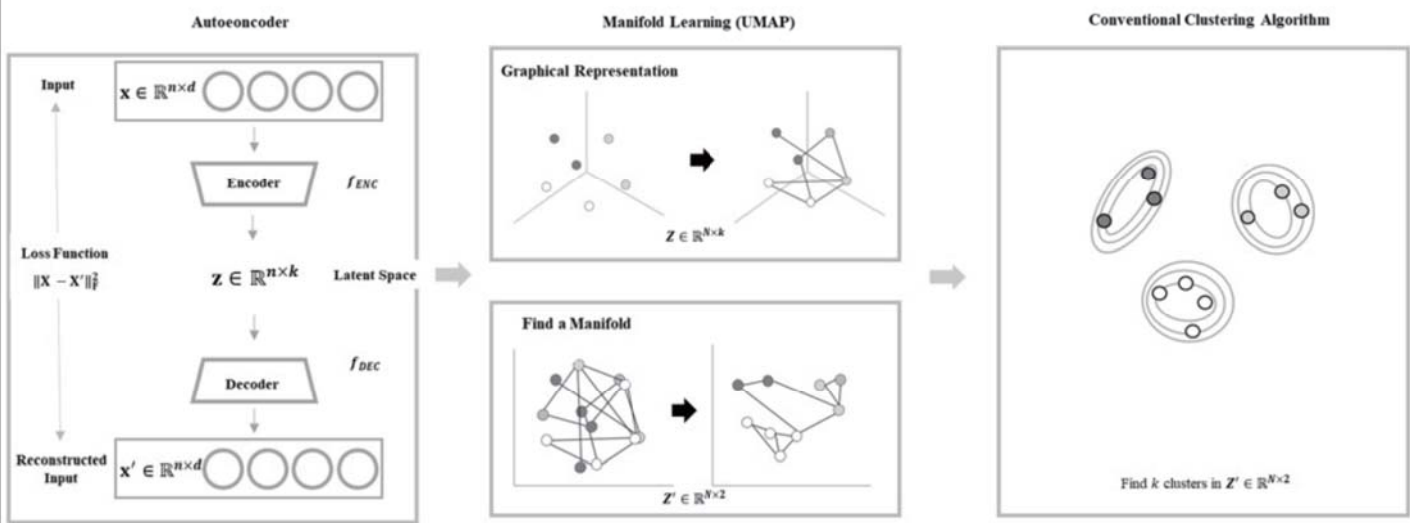
결과 분석



- 모든 데이터를 바탕으로 딥 러닝 기반으로 분류
- 데이터 안의 복잡한 비선형적 관계를 딥러닝을 통해 포착하여 분류 후 그룹 내 heterogeneity가 많이 줄어들 것으로 예상

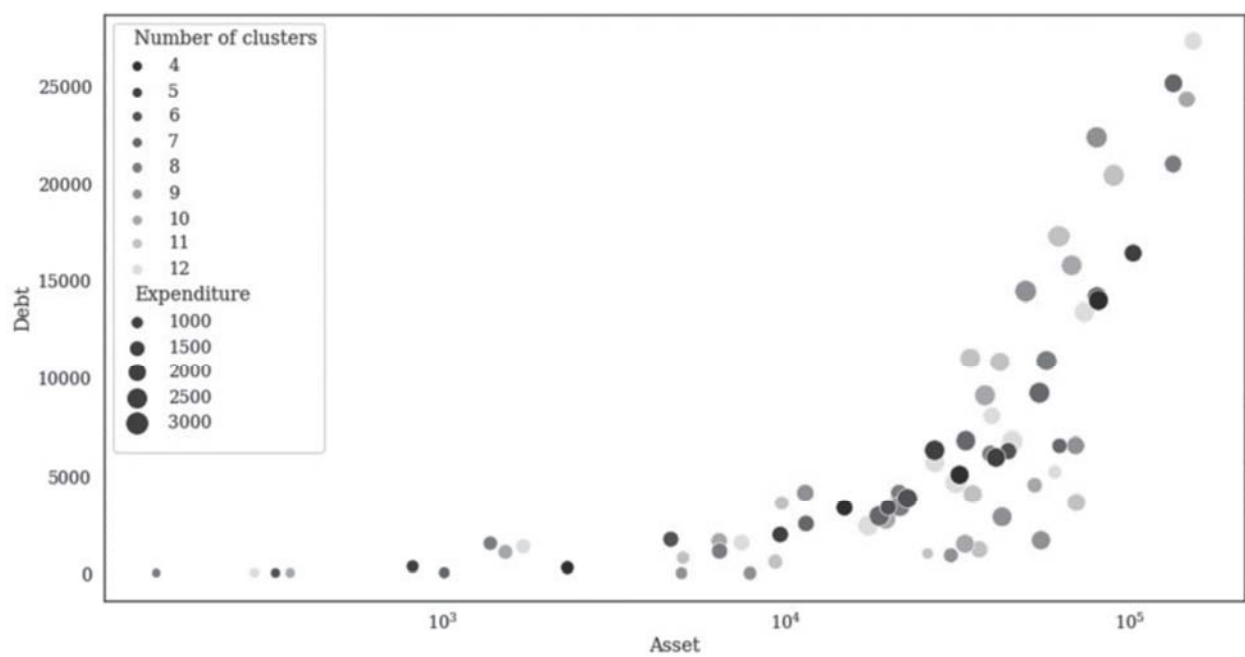
연구 방법

Figure 1 N2D framework for deep clustering by McConville et al. (2021)



분석 결과

Figure 2 Optimal household clusters with different number of clusters



분석 결과

Table 2 Average values (proportions) of asset, debt, expenditure variables of different household clusters

Panel A. Assets							
No.	Deposit savings	Other savings	Long-term rental deposit	Residential housing	Nonresidential real estate	Other real assets	
1	16089.8 (12.1 %)	662.4 (0.5 %)	1343.1 (1.0 %)	42702.6 (32.2 %)	67093.6 (50.6 %)	4673.1 (3.5 %)	
2	10386.6 (18.4 %)	801.0 (1.4 %)	2470.8 (4.4 %)	16426.5 (29.1 %)	22199.5 (39.3 %)	4167.7 (7.4 %)	
3	6507.1 (16.6 %)	353.3 (0.9 %)	1.1 (0.0 %)	30408.0 (77.8 %)	227.3 (0.6 %)	1589.6 (4.1 %)	
4	5189.1 (24.4 %)	219.9 (1.0 %)	82.1 (0.4 %)	12662.5 (59.5 %)	1500.5 (7.1 %)	1612.5 (7.6 %)	
5	5822.6 (28.5 %)	327.1 (1.6 %)	12501.0 (61.3 %)	1.1 (0.0 %)	674.7 (3.3 %)	1076.2 (5.3 %)	
6	2735.1 (43.2 %)	336.3 (5.3 %)	2016.5 (31.8 %)	18.5 (0.3 %)	95.4 (1.5 %)	1132.5 (17.9 %)	
7	469.7 (34.3 %)	33.0 (2.4 %)	474.8 (34.7 %)	31.3 (2.3 %)	18.5 (1.4 %)	339.4 (24.8 %)	
8	62.7 (42.5 %)	1.3 (0.6 %)	78.2 (52.9 %)	0.0 (0.0 %)	0.2 (0.1 %)	5.2 (3.6 %)	
Panel B-1. Debts (Mortgage loans)							
No.	Residential housing	Nonresidential real estate	Long-term rental deposit	Living expense	Business funds	Refinance	
1	2185.7 (19.1 %)	4906.1 (42.9 %)	87.4 (0.8 %)	130.9 (1.1 %)	3208.7 (28.1 %)	123.9 (1.1 %)	
2	1699.0 (26.3 %)	1959.3 (30.3 %)	228.2 (3.5 %)	118.1 (1.8 %)	1382.6 (21.4 %)	108.4 (1.7 %)	
3	4189.9 (81.8 %)	24.2 (0.5 %)	27.5 (0.5 %)	142.9 (2.8 %)	259.5 (5.1 %)	82.4 (1.6 %)	
4	1712.2 (62.2 %)	90.3 (3.3 %)	26.2 (1.0 %)	136.3 (5.0 %)	252.6 (9.2 %)	75.3 (2.7 %)	
5	9.7 (0.3 %)	75.9 (2.2 %)	2456.7 (70.7 %)	56.8 (1.6 %)	62.5 (1.8 %)	16.9 (0.5 %)	
6	0.4 (0.1 %)	3.7 (0.5 %)	146.4 (20.9 %)	60.6 (8.7 %)	74.1 (10.6 %)	14.9 (2.1 %)	
7	7.7 (1.3 %)	0.5 (0.1 %)	8.2 (1.3 %)	15.5 (2.5 %)	18.3 (3.0 %)	5.1 (0.8 %)	
8	0.0 (0.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	
Panel B-2. Debts (Credit loans)							
No.	Residential housing	Nonresidential real estate	Long-term rental deposit	Living expense	Business funds	Refinance	
1	57.1 (0.5 %)	232.3 (2.0 %)	22.7 (0.2 %)	100.5 (0.9 %)	342.8 (3.0 %)	36.1 (0.3 %)	
2	102.9 (1.6 %)	213.4 (3.3 %)	55.0 (0.9 %)	161.6 (2.5 %)	380.0 (5.9 %)	50.9 (0.8 %)	
3	95.9 (1.9 %)	15.5 (0.3 %)	5.8 (0.1 %)	135.4 (2.6 %)	119.2 (2.3 %)	25.4 (0.5 %)	
4	71.6 (2.5 %)	32.9 (1.2 %)	9.9 (0.4 %)	187.6 (6.8 %)	124.8 (4.5 %)	31.4 (1.1 %)	
5	64.4 (1.9 %)	125.4 (3.6 %)	336.1 (9.7 %)	119.5 (3.4 %)	136.5 (3.9 %)	16.0 (0.5 %)	
6	6.0 (0.9 %)	20.1 (2.9 %)	76.7 (11.0 %)	141.7 (20.2 %)	126.9 (18.1 %)	28.5 (4.1 %)	
7	1.6 (0.3 %)	9.0 (1.5 %)	44.8 (7.3 %)	188.5 (30.8 %)	201.2 (32.9 %)	111.8 (18.3 %)	
8	0.0 (0.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	0.6 (100.0 %)	0.0 (0.0 %)	0.0 (0.0 %)	
Panel C. Expenditures							
No.	Foodstuffs	Housing	Education	Medical	Transportation	Communication	Others
1	712.3 (31.4 %)	292.0 (12.9 %)	155.7 (6.9 %)	189.3 (8.4 %)	218.1 (9.6 %)	125.0 (5.5 %)	572.7 (25.3 %)
2	856.1 (27.8 %)	325.7 (10.5 %)	308.3 (12.5 %)	219.1 (7.0 %)	327.3 (10.5 %)	181.7 (5.8 %)	801.4 (25.8 %)
3	708.7 (31.4 %)	257.9 (11.4 %)	230.9 (10.2 %)	164.2 (7.3 %)	217.3 (9.6 %)	142.5 (6.3 %)	537.5 (23.8 %)
4	885.2 (28.5 %)	296.5 (9.5 %)	418.5 (13.5 %)	227.9 (7.3 %)	323.7 (10.4 %)	196.6 (6.3 %)	759.9 (24.4 %)
5	677.6 (31.2 %)	260.5 (12.0 %)	234.3 (10.8 %)	117.5 (5.3 %)	209.0 (9.6 %)	145.5 (6.7 %)	528.3 (24.3 %)
6	604.3 (29.4 %)	368.2 (17.9 %)	165.8 (8.1 %)	116.8 (5.7 %)	207.5 (10.1 %)	147.5 (7.2 %)	447.2 (21.7 %)
7	503.2 (29.5 %)	368.4 (21.6 %)	106.1 (6.2 %)	107.6 (6.3 %)	160.9 (9.4 %)	125.7 (7.4 %)	332.2 (19.5 %)
8	339.7 (36.3 %)	210.5 (22.5 %)	18.4 (2.0 %)	110.0 (11.8 %)	54.1 (5.8 %)	54.1 (5.8 %)	148.5 (15.9 %)

(Unit: KRW 10,000)

분석 결과

Figure 3 Major asset class of households with different level of wealth

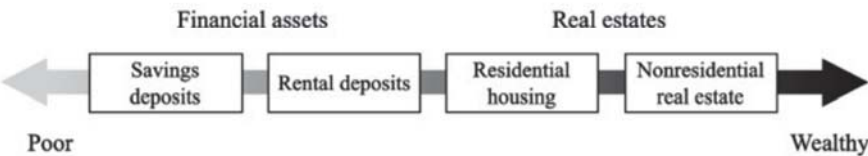
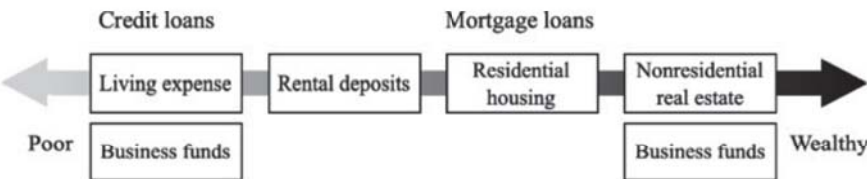
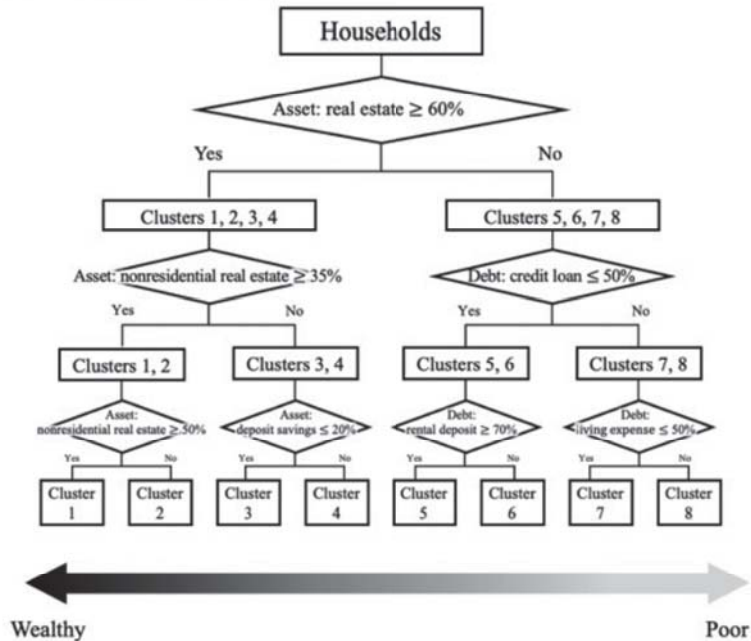


Figure 4 Major loan types of households with different level of wealth



분석 결과

Figure 5 Decision tree for household clusters



분석 결과

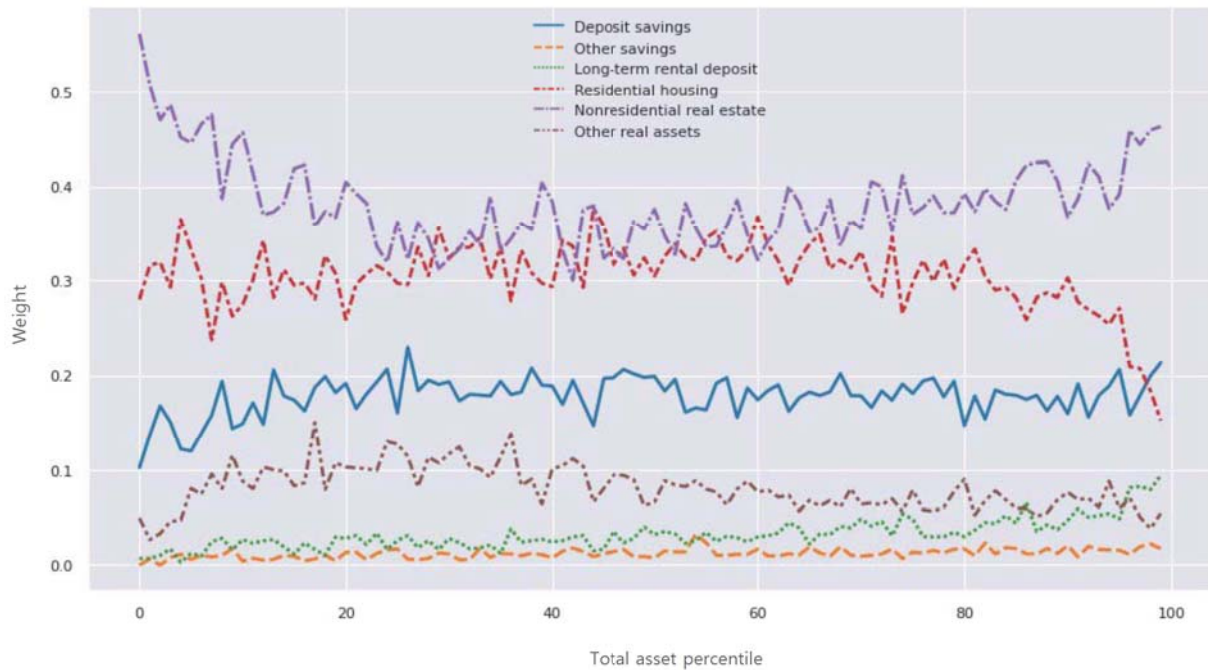
Table 4 Logistic regression results of clusters with respect to socio-demographic variables

Variables	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6		Cluster 7		Cluster 8	
	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio	Coeff	Odds ratio
Constant	-5.224***	0.005	-3.972***	0.019	-2.833***	0.059	-3.025***	0.049	-1.402***	0.246	-0.367***	0.693	-0.088	0.916	-0.107***	0.899
Metropolitan area	0.642***	1.900	-0.646***	0.524	0.622***	1.863	-0.977***	0.376	0.794***	2.212	0.013	1.013	0.021	1.022	-0.205**	0.815
Gender (male)	0.224***	1.251	0.474***	1.607	-0.109***	0.897	-0.014	0.986	-0.411***	0.663	-0.225***	0.799	0.005	1.005	0.137**	1.147
Number of members	-0.349***	0.705	0.062***	1.064	-0.047***	0.954	0.338***	1.402	-0.098***	0.906	0.004	1.004	-0.068***	0.934	-0.525***	0.592
Education (Under middle school)																
High school	0.246***	1.279	-0.064*	0.938	-0.053	0.949	-0.353***	0.702	-0.050	0.951	0.204***	1.227	0.017	1.017	-0.599***	0.549
Higher education	0.773***	2.166	0.129***	1.138	0.134***	1.143	-0.751***	0.472	0.310***	1.363	-0.188***	0.829	-0.435***	0.647	-1.330***	0.264
Home ownership (None)																
Long-term rental	0.489***	1.630	0.597***	1.816	0.063	1.065	-0.463***	0.629	2.197***	8.995	-0.979***	0.376	-2.421***	0.089	-2.414***	0.089
Homeowner	1.486***	4.421	0.656***	1.927	2.806***	16.547	1.648***	5.197	-1.674***	0.187	-2.840***	0.058	-2.911***	0.054	-2.758***	0.063
Age (Under 39)																
40~49	0.665***	1.945	0.596***	1.815	-0.304***	0.738	0.385***	1.470	-0.578***	0.561	-0.170***	0.843	-0.177***	0.838	-0.139	0.870
50~59	1.330***	3.780	1.000***	2.718	-0.584***	0.558	0.164***	1.178	-0.850***	0.427	-0.368***	0.692	-0.251***	0.778	-0.019	0.981
Upper 60	2.539***	12.66	1.235***	3.440	-0.472***	0.624	-0.485***	0.616	-0.957***	0.384	-0.823***	0.439	-0.810***	0.445	-0.168	0.845
Income level (Low-income)																
Mid-income	0.338***	1.403	0.356***	1.428	-0.119***	0.888	-0.031***	0.734	-0.043	0.958	-0.326***	0.722	-0.855***	0.425	-1.753***	0.173
High-income	0.812***	2.252	0.799***	2.224	-0.358***	0.699	-0.257	0.773	-0.331***	0.718	-1.018***	0.361	-2.040***	0.130	-2.299***	0.100
Employment	0.012	1.012	0.628***	1.875	-0.491***	0.612	0.193***	1.213	-0.031	0.970	0.328***	1.388	-0.262***	0.769	-1.091***	0.336
Number of households	5,937		10,644		10,699		10,001		5,614		6,204		4,223		1,598	

*p < .05, **p < .01, ***p < .001

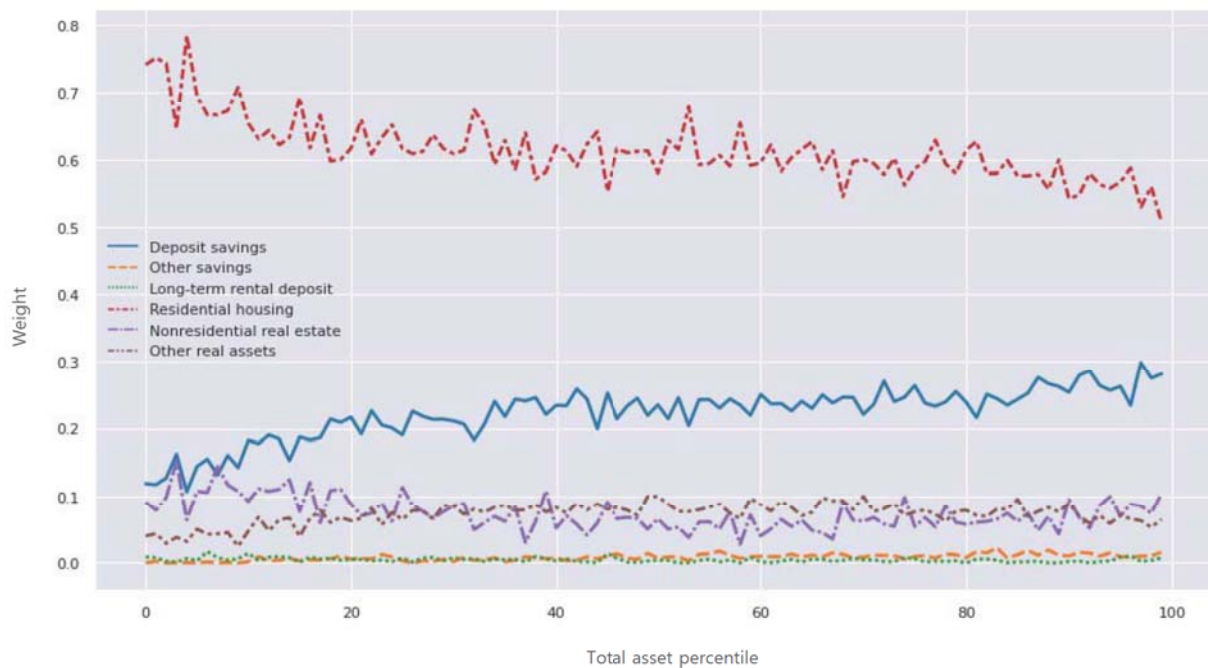
분석 결과

2017~2020 Clustered household portfolio (Label = 2)



분석 결과

2017~2020 Clustered household portfolio (Label = 4)



결론

- 앞서 매우 비선형적이었던 자산 총액과 자산 유형별 비중의 관계가 훨씬 더 선형에 가까운 모습으로 나타나는 것을 확인할 수 있음
- 딥 러닝은 고차원의 복잡한 관계성을 분석하는데 매우 유리함
- 즉, 본 연구에서는 고차원의 복잡한 관계성을 지니는 가구의 portfolio allocation을 딥 러닝을 통해 여러 군집으로 나누어 각 군집 안에서는 선형에 가깝게 나타낼 수 있었다
- 이렇게 실제 데이터를 몇 개의 군집으로 나누어 각각을 선형의 관계성으로 나타낼 수 있다면 consumption-based CAPM도 군집에 따라 따로 모델링하고 이를 통합하여 실제 데이터의 특성을 최대한 잘 반영하는 것도 가능하지 않을까?



Thank you for listening!