THE IMPORTANCE OF THE SOCIO-DEMOGRAPHIC INDICATORS IN THE REGIONAL DISPARITIES IN TURKEY, 1990-1994

42nd European Congress of the Regional Science Association Conference
August 27th-31st, 2002, Dortmund, Germany

Ayse GEDIK
Professor at the Department of City and Regional Planning
Middle East Technical University
06531 Ankara, Turkey
E-mail: gedik@arch.metu.edu.tr; and gedik@metu.edu.tr
Fax: (+90-312-2101250)

Nihan SAHIN
Graduate student at the Department of City and Regional Planning
Middle East Technical University
06531 Ankara, Turkey
E-mail: nisahin@hotmail.com

Sibel SUER
Graduate student at the Department of City and Regional Planning
Middle East Technical University
06531 Ankara, Turkey
E-mail: sibelsuer@hotmail.com

Abstract

The aim of this study is to test the relative importance of the socio-demographic versus economic variables in discriminating the inter-regional differences in Turkey in 1990-94. Our data sources are the various publications of State Institute of Statistics (SIS) and State Planning Organization (SPO).

Univariate descriptive analyses and subsequently factor and logistic regression analyses are carried out when Turkey is divided into two (West and East), and three regions (West, Center, and East). Our findings show that the "socio-demographic" variables, especially those variables related with the female literacy rates and the household size, are the most effective variables rather than the economic variables, in the explanation of the differences between the regions in Turkey. In fact, economic variables are not even found to be statistically significant in our analyses. Subsequently, the trade-off between income per capita and average household size are explored which further supported our previous findings.

Key words: Regional disparity, Socio-demographic variables, Economic variables, Univariate descriptive analysis, Factor and logistic regression analyses.

1. Introduction

Regional economic and social disparities appear as a problem not only in Turkey but also in other developing and developed countries. Therefore, "regional inequality" is an important issue in economic and political agenda of many countries (Shaougang and Angang, 1999, pp. 8-11). In the growth theories in the literature, "the appearance of the human capital theory in the 1960s shifted attention away from capital accumulation to education and training as two of the main constituents of economic growth". Singer in his pioneering article in 1964 , titled as 'International Aid for Economic Development', "he pleaded for more attention to the social aspects of development, opening the path to the social analysis of growth" (Rodriguez-pose, 1998, p.37). Similarly, development economists such as Myrdal (1957) and Kuznets (1966) emphasized the influence of non-economic variables in growth rates.

However, due to the large gulf between theoretical studies and the related empirical observations, the research about the relationship between social factors and growth lost its importance in the early 1970s. Subsequently, in late 1970s, " Fred Hirsch (1977) in his book *Social Limits to Growth* inaugurated a new path for the study of growth from a social perspective. He asserts that social as well as physical resources impose a constraint to growth" (Rodriguez-pose, 1998, p. 38). Likewise, in the 1990s, with the emergence of the concepts like technological development, innovation and learning capacity, some authors such as Romer (1989), Barro (1991), and Young (1991; 1993) paid more attention to human capital issues, especially education and training (Rodriguez-pose, 1998, p. 40). Thus, in later years, in the studies concerning the reduction of regional inequalities, "GDP per capita" as well as the concepts of social and human development (demography, education, culture, health, etc.) gained importance.

On the other hand, in Turkey, there is still a general consensus that the economic indicators are more important than the socio-demographic indicators in terms of regional disparities. Therefore, governments usually, at least up until early 1990's, put emphasis disproportionately more on the economic rather than the socio-demographic factors in their development strategies, investment policies and expenditures.

2. The purpose of the study

There are various studies in Turkey which assess the development level of the provinces using mainly the factor analysis by employing various indicators related to social, demographic, and economic attributes, and levels of various different

infrastructure and services in the provinces (Dincer 1996; Yener and Koc 2001). However, the purpose of this study is to test the relative importance of the socio-demographic variables versus economic variables in the explanation of the inter-regional differences in Turkey.

Another point is that in this study, the independent variables which are included in the analyses are people- (rather than spatial) oriented. For example, instead of level and quality of the infrastructure and services related to health or education, their outcome/effect on the people of the province, such as the education level of the people or the infant mortality rates are considered.

Detailed various univariate descriptive tests, as well as the factor and logistic regression analysis are carried out when Turkey is divided into seven, three, and two regions. The data source is various publications of State Institute of Statistics (SIS) and State Planning Organization (SPO).

3. Data and Methodology

For this study, eleven socio-demographic, and six economic variables, i.e., total of 17 independent variables are selected (see Table 1). The data is for the years 1990, and 1994, and for the 73 provinces. However, the two provinces of the East and the Southeast Turkey, i.e., Hakkari and Sirnak are excluded from the analysis in order not to distort our findings because they were outliers especially in terms of the socio-demographic variables.

Firstly, the univariate descriptive analyses are carried out for each of the seven regions of Turkey. Subsequently, the analyses are carried out for the three regions and two regions. The three regions are as follows: (1) "Western" Turkey (Marmara, Aegean, and Mediterranean regions); (2) "Central" Turkey (Central Anatolia and Black Sea regions); and (3) "Eastern" Turkey (Eastern and South Eastern Anatolia regions). The Western and Eastern Turkey are considered to be the most and the least developed regions, respectively; whereas Central Anatolia has the average values for the country as a whole, according to our descriptive tests and to various other studies (Dincer et. al. 1996; Yener and Koc 2001, pp.143-45). The two regions are: (1) "Eastern" Turkey (Eastern and South Eastern regions); and (2) the rest of five regions in Turkey which is called in this study as "Western" Turkey.

The detailed univariate analyses are applied in order to obtain transparency in the subsequent multivariate tests. These univariate tests are in terms of: (1) the study of average values of our variables for the seven, three and two regions, and the inter-

regional differences in these values; (2) normality tests; and (3) the tests of equality of group means (one-way analysis of variance).

Subsequently, the multivariate tests such as factor and logistic regression analyses are applied on two and three regions. The tests could not be carried out for seven regions because of small sample size which could be as low as 7-8 in three of the regions, i.e., Aegean, Mediterranean, and Southeastern Turkey (see Table 1).

Table 1. The list of variables

	No:	Symbol:	Variable:
Socio-demographic variables	1	TFR[1]	Total fertility rate, 1990
	2	AHS[1]	Average household size,1990
	3	AGE0-14[2]	Percentage of 0-4 age population in total population (%), 1990
	4	E(0)[3]	Life expectancy at birth, 1996
	5	IMR[2]	Infant mortality ratio (‰), 1990
	6	LTR[2]	Literacy ratio (%), 1990
	7	F_LTR[2]	Female literacy ratio (%), 1990
	8	SSCR[2]	Schooling ratio in secondary school (%), 1994-95
	9	F_SSCR[2]	Female schooling ratio in secondary school (%), 1994-95
	10	HSCR[2]	Schooling ratio in high school (%), 1994-95
	11	F_HSCR[2]	Female schooling ratio in high school (%), 1994-95
Economic variables	12	GDP[4]	GDP per capita (at 1987 prices, million TL), 1990
	13	NAGR[2]	Percentage of non-agricultural active population in total last week's economic active population (%), 1990
	14	F_NAGR[2]	Percentage of non-agricultural active female population in total last week's economic active female population (%), 1990
	15	AUTO[1]	Number of automobiles per 10,000 population, 1994
	16	ELEC[1]	Electricity consumption per person (KWh), 1994
	17	TEL[1]	Consumption of telephone communication ("kontur") (TL), 1994

Source: (1) DPT, 1996, p:87; (2)DIE, 2001; (3) UNDP, 1998; (4) DPT, 2001

4. Univariate descriptive analyses

The arithmetic averages of the variables indicate lesser development levels as we move from West to East (see Tables 2 and 3). As previously stated, Central Turkey and Black Sea regions have medium level of development.

For example, moving from the western to eastern regions, "total fertility rate" and "average household size" increase while variables concerning education such as "literacy rate" and "schooling ratios" decrease dramatically. Underdevelopment of the East in socio-demographic variables is especially startling for female education. For example, "female schooling ratio in high school" in the West is at least three times of

Table 2. Arithmetic averages of variables

	Variable	With 7 regions							With 3 regions			With 2 regions		National (TURKEY)
		Marmara	Aegean	Mediter.	Central	Black Sea	East	S.East	West	Central	East	West	East	
Socio-demographic variables	TFR	**2.09**	**2.22**	**2.53**	**2.59**	**2.70**	4.01	4.99	**2.25**	**2.65**	4.40	**2.46**	4.40	2.95
	AHS	**4.24**	**4.62**	**4.74**	**5.26**	5.61	6.91	6.98	**4.50**	**5.45**	6.94	**5.00**	6.94	5.49
	AGE0_14	**28.02**	**30.79**	**34.88**	**35.71**	**34.74**	43.74	48.02	**30.83**	**35.19**	45.41	**33.13**	45.41	36.24
	E(0)	**70.59**	**71.99**	66.07	64.73	65.85	61.58	65.53	**69.77**	65.33	63.12	**67.43**	63.12	66.33
	IMR	**60.50**	**65.88**	**58.86**	68.85	71.47	77.64	66.43	**61.76**	70.25	73.28	**66.25**	73.28	68.00
	LTR	**86.22**	**82.79**	**81.64**	**81.96**	**78.20**	67.62	59.92	**83.84**	**79.95**	64.62	**81.78**	64.62	77.43
	F_LTR	**80.17**	**75.24**	**73.53**	**74.26**	**69.53**	54.36	43.92	**76.73**	**71.73**	50.30	**74.09**	50.30	68.06
	SSCR	**67.88**	**54.95**	**53.49**	**56.46**	50.17	36.55	28.35	**59.71**	**53.09**	33.36	**56.21**	33.36	50.42
	F_SSCR	**59.79**	**47.06**	**48.41**	**46.17**	39.12	24.78	16.70	**52.53**	**42.39**	21.64	**47.17**	21.64	40.70
	HSCR	**47.81**	**36.69**	**36.68**	**38.20**	**35.44**	24.22	18.03	**41.13**	**36.72**	21.82	**38.80**	21.82	34.50
	F_HSCR	**43.30**	**31.88**	**32.29**	**30.56**	26.13	15.19	9.75	**36.56**	**28.19**	13.07	**32.14**	13.07	27.30
Economic variables	GDP	**2.09**	**1.57**	**1.38**	1.11	0.95	0.60	0.81	**1.72**	1.03	0.68	**1.35**	0.68	1.18
	NAGR	**51.97**	**40.52**	**41.50**	**38.72**	27.94	27.34	30.64	**45.38**	32.95	28.62	**38.81**	28.62	36.23
	F_NAGR	**24.96**	**17.24**	**16.18**	**15.11**	7.11	6.18	5.48	**20.03**	10.82	5.91	**15.17**	5.91	12.82
	AUTO	**472.40**	**492.25**	**428.86**	**381.54**	283.33	167.36	153.43	**466.56**	328.93	161.94	**393.85**	161.94	335.06
	ELEC	**1695.90**	**966.00**	**956.00**	703.69	621.13	364.27	457.57	**1255.04**	659.46	400.56	**940.40**	400.56	803.54
	TEL	**945.20**	**746.63**	**654.71**	**611.23**	540.73	422.64	360.71	**800.32**	573.46	398.56	**680.47**	398.56	609.00
	n	10	8	7	13	15	11	7	25	28	18	53	18	71

Notes: For the definitons of variables see Table1. Those values which are relatively better than the national average are indicated in bold.

Table 3. Arithmetic averages of standardized values (z-score) of the variables

		With 7 regions							With 3 regions			With 2 regions	
	Variable	Marmara	Aegean	Mediter.	Central Anatolia	Black Sea	Eastern Anatolia	S.Eastern Anatolia	Western Anatolia	Central Anatolia	Eastern Anatolia	Western Anatolia	Eastern Anatolia
Socio-demographic variables	TFRx	**0.78**	**0.66**	**0.38**	**0.33**	**0.23**	-0.95	-1.84	**0.63**	**0.27**	-1.30	**0.44**	-1.30
	AHSx	**1.08**	**0.76**	**0.65**	**0.20**	-0.11	-1.23	-1.29	**0.86**	**0.04**	-1.25	**0.43**	-1.25
	AGE0_14x	**1.14**	**0.75**	**0.19**	**0.07**	**0.21**	-1.04	-1.63	**0.75**	**0.15**	-1.27	**0.43**	-1.27
	E(0)	**1.03**	**1.37**	-0.06	-0.39	-0.12	-1.15	-0.20	**0.84**	-0.24	-0.78	**0.27**	-0.78
	IMRx	**0.58**	**0.16**	**0.70**	-0.06	-0.26	-0.73	0.12	**0.48**	-0.17	-0.40	**0.14**	-0.40
	LTR	**0.91**	**0.56**	**0.44**	**0.47**	**0.08**	-1.02	-1.82	**0.67**	**0.26**	-1.33	**0.45**	-1.33
	F_LTR	**0.93**	**0.55**	**0.42**	**0.47**	**0.11**	-1.05	-1.84	**0.66**	**0.28**	-1.36	**0.46**	-1.36
	SSCR	**1.11**	**0.29**	**0.20**	**0.38**	-0.02	-0.88	-1.40	**0.59**	**0.17**	-1.09	**0.37**	-1.09
	F_SSCR	**1.09**	**0.36**	**0.44**	**0.31**	-0.09	-0.91	-1.37	**0.67**	**0.10**	-1.09	**0.37**	-1.09
	HSCR	**1.10**	**0.18**	**0.18**	**0.31**	**0.08**	-0.85	-1.36	**0.55**	**0.18**	-1.05	**0.36**	-1.05
	F_HSCR	**1.18**	**0.34**	**0.37**	**0.24**	-0.09	-0.89	-1.30	**0.68**	**0.07**	-1.05	**0.36**	-1.05
Economic variables	GDP	**1.42**	**0.60**	**0.30**	-0.11	-0.37	-0.92	-0.58	**0.84**	-0.25	-0.79	**0.27**	-0.79
	NAGR	**1.11**	**0.30**	**0.37**	**0.18**	-0.58	-0.63	-0.39	**0.64**	-0.23	-0.54	**0.18**	-0.54
	F_NAGR	**0.95**	**0.35**	**0.26**	**0.18**	-0.45	-0.52	-0.57	**0.56**	-0.16	-0.54	**0.18**	-0.54
	AUTO	**0.71**	**0.81**	**0.48**	**0.24**	-0.27	-0.87	-0.94	**0.68**	-0.03	-0.89	**0.30**	-0.89
	ELEC	**1.46**	**0.27**	**0.25**	-0.16	-0.30	-0.72	-0.57	**0.74**	-0.24	-0.66	**0.22**	-0.66
	TEL	**1.12**	**0.46**	**0.15**	**0.01**	-0.23	-0.62	-0.83	**0.64**	-0.12	-0.70	**0.24**	-0.70
SOC-DEM		0.99	0.54	0.35	0.21	0.00	-0.97	-1.27	0.67	0.10	-1.09	0.37	-1.09
ECON		1.13	0.46	0.30	0.06	-0.37	-0.71	-0.65	0.68	-0.17	-0.69	0.23	-0.69
n		10	8	7	13	15	11	7	25	28	18	53	18

Notes: The signs of TFRx, AHSx, AGEx and IMRx are inverted in order to make these variables consistent with the others. Those values which are relatively better than the national average are indicated in bold.

that in the East. On the other hand, regional averages for the economic variables decrease as we move from west to east.

These values indicate to us as if there is a trade-off between demographic variables, and the economic well-being. For example, "total fertility rate" and "average household size" in the East is 2.1 and 1.6 times of that in the West, respectively; whereas, "GDP p.c." and the "ratio of non-agricultural active population" in the West is 2.6 and 1.3 times of that in the East.

The East and Southeast are the only regions where the values of all of the "socio-demographic" variables are lower than the average values for Turkey. On the other hand, Black Sea region (and Central Turkey, at least for some of the economic variables) as well as the East and Southeast have lower than the average values for all of the "economic" variables (see Tables 1 and 2).

It is interesting to note that, in terms of standardized values (see Table 2 and Figures 1-3): (1) the Western regions with higher than average values for Turkey, have almost equal position on socio-demographic and economic variables with almost perfect positive correlation from Mediterranean to Marmara regions; (2) the Central and Black Sea regions have almost average values of Turkey on both types of variables, with economic variables being slightly lower than the national average and vice versa; (3) and the East and Southeast, although they have lower than the average values for both types of variables, they fare better on the economic variables, and more so in the Southeast[1].

Figure 1. Averages of the standardized values of the socio-demographic variables and economic variables in terms of 7 regions
Source. Table 3

Figure 2. Average of the standardized of the socio-demographic variables and economic variables in terms of 3 regions
Source. Table 3

Figure 3. Average of the standardized of the socio-demographic variables and economic variables in terms of 2 regions
Source. Table 3

In the two regions analyses, since in the category of "West", Central Anatolia and the Black Sea regions are also included, the differences between West and East are attenuated, especially in the economic variables.

In the, Kolmogorov-Smirnoff normality tests, total fertility rate (TFR), percentage of population in the age group of 0-14 (AGE0_14), percentage of non-agricultural active population (NAGR, and F_NAGR), consumption of electricity (ELEC) and telephone communication (TEL) are found to be positively, and literacy rates (LTR and F_LTR) to be negatively skewed (see Table 4).

In the test of equality of group means (one-way analysis of variance), the independent variables are considered individually (see Table 5). Non-parametric tests are carried out because of the non-homogeneous variances between regions for most of the variables. In all three types of regionalization, the two socio-demographic variables, i.e., the female literacy rate (F_LTR) and the average household size (AHS), have the highest values[2] than any one of the economic variables, including GDP p.c. (see Table 5). Values of the economic variables rank after most of the socio-demographic variables. The later all refer to past and/or present fertility, and education. The infant mortality rate (IMR) is the least important in all three types of regionalization. This indicates the success of the nation-wide campaign of inoculation of infants beginning with 1990's in Turkey-- as well as the relative ease of improving indices related to health and mortality relative to the fertility and education of females which are at least partly related to the cultural norms in the different regions of the country.

Table 4. Tests of normality

		Not in logs			In logs		
		Kolmogorav-Smirnov [a]			Kolmogorav-Smirnov [a]		
	Variable	Statistic	d.f.	Sig.	Statistic	d.f.	Sig.
Socio-demographic variables	TFR*	0.202	71	0.000	0.161	71	0.000
	AHS	0.082	71	0.200			
	AGE0_14**	0.130	71	0.005	0.093	71	0.200
	E(0)	0.780	71	0.200			
	IMR	0.111	71	0.031			
	LTR*	0.167	71	0.000	0.201	71	0.000
	F_LTR*	0.192	71	0.000	0.236	71	0.000
	SSCR	0.069	71	0.200			
	F_SSCR	0.068	71	0.200			
	HSCR	0.071	71	0.200			
	F_HSCR	0.078	71	0.200			
Economic variables	GDP	0.113	71	0.026			
	NAGR**	0.150	71	0.000	0.108	71	0.040
	F_NAGR**	0.222	71	0.000	0.099	71	0.080
	AUTO	0.109	71	0.037			
	ELEC**	0.148	71	0.001	0.063	71	0.200
	TEL**	0.150	71	0.000	0.070	71	0.200

Notes*:* (1) a: Lilliefors significance correction; (2) * not normal even in logs; ** normal in logs.

In summary, when the variables are considered individually, socio-demographic variables differ between regions much more than the economic variables whether the country is divided into seven, three or two regions. Economic variables differ even less between regions when the country is divided only into two regions (as East and West) (see Table 5). This is because of the fact that in the two-region analysis, in the West, Central and the Black Sea regions are also included whose economic development are below the average of Turkey (see Tables 2 and 3).

5. Multivariate Tests: Factor and Logistic Regression Analyses

Firstly, in order to encounter the problem of multicollinearity in the logistic regression, factor analysis is carried out in order to obtain two uncorrelated distinct factors one for socio-demographic, and the other for the economic variables. Secondly the logistic regression is applied.

Table 5. Univariate tests: One-way analysis of variance

	7 regions				3 regions				2 regions			
		Krusgal-Wallis test				Krusgal-Wallis test				Man-Whitney test		
	Variable	Chi-square	d.f.	Sig.	Variable	Chi-square	d.f.	Sig.	Variable	U-statistic	d.f.	Sig.
Socio-demographic variables	F_LTR	48.818	6	0.000	AHS	44.400	2	0.000	F_LTR	32.0	1	0.000
	LTR	48.443	6	0.000	F_LTR	42.145	2	0.000	LTR	39.0	1	0.000
	AHS	46.876	6	0.000	LTR	41.234	2	0.000	AHS	49.5	1	0.000
	AGE0_14	46.180	6	0.000	AGE0_14	39.899	2	0.000	AGE0_14	54.0	1	0.000
	TFR	41.599	6	0.000	TFR	36.890	2	0.000	F_HSCR	80.0	1	0.000
	E(0)	41.481	6	0.000	F_HSCR	34.380	2	0.000	TFR	85.0	1	0.000
	F_HSCR	39.097	6	0.000	F_SSCR	32.423	2	0.000	F_SSCR	87.0	1	0.000
	F_SSCR	36.459	6	0.000	E(0)	29.642	2	0.000	SSCR	97.0	1	0.000
	HSCR	34.842	6	0.000	SSCR	28.792	2	0.000	HSCR	99.5	1	0.000
	SSCR	34.796	6	0.000	HSCR	27.561	2	0.000	E(0)	165.5	1	0.000
	IMR	12.339	6	0.054	IMR	9.530	2	0.009	IMR	360.5	1	0.123
Economic variables	AUTO	39.711	6	0.000	AUTO	38.970	2	0.000	AUTO	82.0	1	0.000
	F_NAGR	38.683	6	0.000	GDP	35.555	2	0.000	GDP	139.0	1	0.000
	GDP	38.234	6	0.000	F_NAGR	34.620	2	0.000	F_NAGR	141.0	1	0.000
	ELEC	37.780	6	0.000	ELEC	32.768	2	0.000	ELEC	142.5	1	0.000
	NAGR	32.751	6	0.000	TEL	27.034	2	0.000	TEL	149.0	1	0.000
	TEL	32.295	6	0.000	NAGR	26.406	2	0.000	NAGR	250.0	1	0.003

5.1. Factor analysis. All of the 17 independent variables are entered into the principal component analysis in two alternative ways: one in which none of the variables are in logs; the second in which some of the variables which were found to be non-normal are

in terms of logs (see Tables 4 and 6). “Factor analysis is a statistical technique used to identify a relatively small number of factors that can be used to represent relationships among sets of many interrelated variables” (Norusis, 1994, p. 47). In this study, however, all of the 17 variables were heavily loaded only on one factor, and thus it was impossible to obtain two different factors, as we initially hoped (see Table 6). Since the second component explained less than 10% of the total variance, factor score coefficients are shown only for the first component. The total variance, which was accounted by these two components, was about 80%.

Table 6. Factor score coefficient matrix
(Principal component)

No variable is in log		Some variables are in log	
F_HSCR *	0.080	F_SSCR *	0.76
SSCR *	0.079	F_HSCR *	0.76
F_SSCR *	0.079	SSCR *	0.75
LTR *	0.077	HSCR *	0.74
HSCR *	0.077	L_TFR *	-0.73
AHS *	-0.076	AHS *	-0.73
F_LTR *	0.076	L_AGE *	-0.72
AGE *	-0.075	L_LTR *	0.72
TFR *	-0.073	L_TEL	0.72
AUTO	0.073	L_F_LTR *	0.70
TEL	0.071	AUTO	0.69
GDP	0.069	L_F_NAGR	0.68
NAGR	0.067	L_NAGR	0.67
F_NAGR	0.059	GDP	0.65
ELEC	0.058	L_ELEC	0.66
E(o) *	0.050	E(o) *	0.47
IMR *	-0.044	IMR *	-0.42
KMO	0.885	KMO	0.90
Total cum.variance explained (%)	79.254	Total cum. variance explained (%)	80.522

Notes: Socio-demographic variables are indicated with asterisk.

Although we were unable to obtain two distinct meaningful components, the results were insightful and were consistent with the previous analyses. If we consider that these total of 17 socio-demographic-economic variables represent level of development of the 71 provinces in Turkey, then when the factor score coefficients are ranked on the first component, we see that, all but except the two (i.e., the expected life at birth, E (0), and the infant mortality rate, IMR), socio-demographic variables load most heavily. All the economic variables have lesser coefficients. For example, all the economic variables

including GDP per capita (GDP) have smaller coefficients. Although this is counter-intuitive, it is consistent with some other studies in Turkey (Yener 2001, p. 143).

5.2. Logistic regression analyses. As previously stated, because of the small sample size in some regions, the case when Turkey is divided into seven regions is not included in the analyses. Thus only the two (West and East) and region (West, Center, and East) models are analyzed.

The analysis is carried in two methods. Firstly, the selection of the variables are based on the theoretical considerations. Total of three variables are selected: one variable for each of the three indicators concerning fertility, education, and the economy of the provinces. The selection is based on statistical as well as theoretical concerns.

For example, among variables about education, indicators for females and for both sexes are very highly correlated (correlation coefficient being larger than 0.95). Therefore, considering such a high correlation, besides our previous findings in the univariate descriptive tests, only the indices about female education are included in the logistic regression. The tests for female education are repeated for literacy (F_LTR), schooling ratios for secondary school (F_SSCR) and for high school education(F_HSCR). The results of the later two variables is not presented, because in every test, female literacy rate (F_LTR) performed much better than the other two variables. The schooling ratio for the primary school edulation is not included because of its inconsistent values[3].

Among the indices concerning fertility, average household size (AHS), besides its high performance in the univariate tests, is thought to be better representative (than the other two variables, e.g.,TFR, and Age0_14) since it takes account of the consequences of the past and present fertility on the household. Furthermore, it enabled us to carry out explorative comparative analysis between the effects of income per capita and the average household size.

As for the economic variables, the variables concerning the percentage of non-agricultural active population (NAGR and F_NAGR) are not included in the analysis. This is because of the fact that some provinces which were heavily agricultural oriented (such as Aydin, Antalya, etc in the West) have very high level of economic development, which resulted in inconsistent results in our analyses. The rest of the four economic variables, since they performed poorly in the previous univariate tests, they were included into the analyses in various ways: all together (GDP, AUTO, ELEC, and TEL); or one by one; or together with the socio-demographic variables together.

Secondly purely statistical methods are applied in selecting variables for the analysis. Initially, all of the 17 variables are entered together. Those variables which are statistically significant are re-entered into the analysis for calibration. If none of them are significant, then stepwise method is used; and those which are found to be significant, are re-entered into the analyses.

As we shall see below in detail, in both of the approaches, the best models are those with the varibles of the female literacy rate (F_LTR) and the average household size (AHS). None of the economic variables were found to be statistically significant and/or performed well.

5.2. 1. Logistic regression analysis with two regions. Firstly, the average household size (AHS), the female literacy rate (F_LTR), and the economic variables, together or one by one, are entered into the analysis. None of the economic variables were statistically significant; and their significance levels for Wald statistic ranged between 0.403 and 0.898. On the other hand, significance level of the other two variables (average household size (AHS) and and female literacy rate (F_LTR)) were 0.122 and 0.016, respectively (see Model 3 in Table 7). Only two provinces are misclassified.

Table 7. Estimated values of odds ratios (Exp (B)) in the logistic regression analysis: two regions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Variables	Stepwise [1]	Stepwise [2]	Enter AHS, F_LTR	Stepwise [3]	Enter L_GDP	Enter Income_pc	Enter Income_pc_	Enter Income_pc_ and F_LTR
AHS		31.857**	6.813					
F_LTR	0.652**		0.656*					0.623**
GDP					0.001**			
AUTO				0.982**				
INCOME_pc						0.259*		
INCOME_pc_							0.825**	1.114
Constant	7.0E+11**	0.000**	7137823	27.440*	7.9E+18**	1.4E+13*	65.438*	7.2E+11*
- 2LL	25.03	30.956	21.539	45.212	55.267	15.866	66.098	23.423
Pseudo R^2(Nagelkerke)	0.8	0.74	0.83	0.58	0.44	0.88	0.27	0.814
Correct classification (%)	91.5	93	97.2	85.9	87.3	94.4	83.1	93
No of misclassified provinces	6	5	2	10	9	4	12	5
c (ROC)	0.966	0.948	0.97	0.914	0.854	0.988	0.774	0.974

Notes: Asterisks indicate significance levels of the Wald statistics: *significant at 0.05 level; **significant at 0.01 level. Those variables which are found to be significant in the stepwise method, are subsequently re-entered into the analysis. (1) Stepwise with all 12 variables; 10 variables (exclude F_SSCR and F_HSCR). (2) Stepwise with 6 variables (include AHS, F_LTR, and GDP-TEL). (3) Stepwise with 4 economic variables (GDP-TEL).

Furthermore, explorative analyses are carried out with only one of the variables in each logistic regression analysis. The average household size (AHS) and the female literacy rate (F_LTR) performed equally well (see Models 1-2 in Table 7); whereas, only the number of automobiles per 10,000 population (AUTO) and the log of GDP per capita (L_GDP) were statistically significant, but both performed very poorly in terms of various indices used for the "goodness of fit"(see Models 4-5 in Table 7).

Secondly, stepwise method is applied with 12, 10, and 6 variables: either only the female literacy rate (F_LTR), or only the average household size (AHS) was statistically significant, respectively (see Models 1-2 in Table 7).

Among the alternative models, model 3 performs the best. Only the two provinces in the East (Erzincan and Malatya) are mis-classified. The odds of a province to be in the East increasee by a factor of 6.813 if the average household size increases by one person; and by a factor of 0.656 if the female literacy rate is increase by 1%--where odds equals to the ratio of the probability a province to be located in the East to the probability to be located in the West (Odds=Prob (East)/Prob (West)).

B	Wald sign.	Odds ratio	95% Confid. interv. (lower and upper bounds)
AHS	1.919	0.122	6.813 0.579 and 77.746
F_LTR	-42.135	0.016	0.656 0.465 and 0.925
Constant	15.781	0.231	7137823

5.2. 2. Logistic regression analysis with three regions. The East is considered as the reference category. Firstly, like in the analyses for two regions, firstly, the average household size (AHS), the female literacy rate (F_LTR), and the economic variables, together or one by one, are entered into the analysis. None of the economic variables were significant either in terms of likelihood ratio tests (-2LL) or in terms of the Wald statistics; and they all had coefficients of about zero with the respective odds ratios of 1.0. On the other hand, the two variables were significant in terms of likelihood ratio tests (-2LL); and their significance levels for the Wald statistic for the average household size (AHS) and and female literacy rate (F_LTR) were 0.009 and 0.027 in the West, and 0.285 and 0.015 in the Central T., respectively (see Model 2 in Table 8). Nine provinces are misclassified.

Secondly, stepwise method is applied with 12, 10, and 7 variables. In all three runs, the average household size (AHS), the female literacy rate (F_LTR), and expected life at birth (E(0)) were all statistically significant in terms of likelihood ratio test (-2LL); and their significance levels for the Wald statistic were 0.003, 0.153, and 0.021 in the West; and 0.211; 0.446, and 0.029 in the Center, respectively (see Model 1 in Table 8). Eleven provinces are misclassified.

Table.8 Estimated Values of odds ratios (Exp(B)) in the logistic regression analysis: three regions

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	Enter [1]		Enter AHS, F_LTR		Enter AHS		Enter F_LTR		Enter INCOME_pc		Enter INCOME_pc_		Enter INCOME_pc_ and F_LTR	
	Odds ratio Exp(B)		Odds ratio Exp(B)		Odds ratio Exp(B)		Odds ratio Exp(B)		Odds ratio Exp(B)		Odds ratio Exp(B)		Odds ratio Exp(B)	
Variables	West	Center	West	Center	West	Center	West	Center	West	Center	West	Center	West	Center
AHS	**0.0107**	**0.201	**0.019**	**0.233	**0.004**	**0.004**								
E(0)	**1.191*	**1.195												
F_LTR	**1.340	**1.523*	**1.521*	**1.535*			1.789**	1.478**					1.460**	1.689**
INCOME_pc									**5.444**	**3.763*				
INCOME_pc_											**1.523**	**1.121	0.997	0.590**
2LL	56944.000		69565.000		76461.000		86.885		53655.000		106472.000		56.348	
Pseudo R^2 (Nagelkerke)	0.841		0.784		0.734		0.689		0.854		0.549		0.843	
Correct Classification (%)	84.5		87.3		81.7		73.2		84.5		73.2		81.7	
No of mis-classified provinces	11		9		13		19		11		19		13	

Notes: Asterisks indicate significance levels: left side of the odds ratios, refer to likelihood ratio tests (chi-square), and right side to the Wald statistics. ***** significant at 0.05 level. ****** significant at 0.01 level. Those variables which are found to be significant, are subsequently re-entered into the analysis. (1) These three variables were always significant in the likelihood ratio tests (chi-square) with all 12 variables; 10 variables (exclude F_SSRC and F_HSCR); or 7 variables (include only AHS, E(0), F_LTR, GDP-TEL).

Subsequent to the stepwise method, each of the three above stated variables are entered individually. Among these three runs, the one with the average household size (AHS) performed the best (see Model 3 in Table 8). In terms of goodness of fit, it is similar to the Model 1 in which all three variables (the average household size (AHS), the female literacy rate (F_LTR), and expected life at birth (E(0)) are entered into the analysis. The number of misclassified provinces increased from 11 to only 13.

Among the alternative models, Model 2 performs the best. Nine provinces are misclassified between adjacent regions: 5 in the West (predicted in the Central T.), 2 in the Central T. (predicted in the West), and 2 in the East (predicted in the Central T.). The odds of a province to be in the West rather than the East increase by a factor of 0.019 if the average househould size (AHS) increases by one member; and by a factor of 1.521 if the female literacy rate (F_LTR) increases by 1%. Similarly, the odds for a province to be in the Central T. rather that the East increases by a factor of 0.233 if the average household size increases by one member; and by a factor of 1.535 if the female literacy rate (F_LTR) increases 1%. In other words, the odds ratios for the female literacy rate (F_LTR) for the West and Central is very similar to each other. However, the average household size (AHS) has larger effect for the West than the Central.

	B	Wald sign.	Odds ratio	95% Confid. interv. (lower and upper bounds)
(West)				
AHS	-3.924	0.009	0.01976	0.00105 and 0.370
F_LTR	0.419	0.027	1.521	1.048 and 2.206
Constant	-6.278	0.688		
(Central)				
AHS	-1.455	0.285	0.233	0.01624 and 3.355
F_LTR	0.429	0.015	1.535	1.086 and 2.169
Constant	-19.145	0.174		

5.2.3. Logistic regression with income per capita

In the previous analyses with two and three regions, the GDP per capita, as well as other economic variables, were not statistically significant and/or performed poorly—whether they were entered together with other variables or by themselves, into the analyses. In this section, income per capita is tested by logistic regression. The analyses are explorative in nature, and the findings are tentative, and should be considered with due caution.

The accuracy of the data obtained by interview about the personal or household "incomes" is usually doubtful. Furthermore, the data used in this study is obtained indirectly by dividing the average household income of the regions by the average household size of the respective provinces[4].

In the logistic regression analyses the goodness of fit of the models with INCOME per capita (INCOME_pc) was similar and even better than the models with the average household size (AHS), both in two and three region analyses (see Models 6 and 2 in Table 7; and see Models 3 and 5 in Table 8).

Subsequently, the income per capita is adjusted such that the average household size (AHS) of the provinces in the East (i.e.,East and Southeast) are decreased by the factor of 2/7 ((AHS-(2/7)*(AHS)) such that the average household size in the East (7 members) is now equal to the average in the other regions of Turkey (5 members) (see Table 2). This adjustment resulted in approximately 40% increase in the per capita incomes in the East (from 17.53 to 24.55).

	West	Center	East	West	East	Turkey
GDP	1.72	1.03	0.68	1.68	0.68	1.18
INCOME_pc	41.92	26.17	17.53	33.60	17.53	29.53
INCOME_pc_	41.92	26.17	24.55	33.60	24.55	31.30

In subsequent analyses, the "adjusted income per capita" performed poorly, relative to the case when only the female literacy rate was included in the analysis (see Models 1 and 7 in Table 7; and Models 4 and 6 in Table 8). Likewise, when it was included in the analysis to gether with female literacy rates, it was either not significant, or its odds ratio was near 1.0, or its odds ratio was even less than one for the Central T. The later meant that in the logistic regression with three regions, Central T. (Central and Black Sea regions) fared worse than East Turkey when the per capita incomes were adjusted for equal household size (see Model 8 in Table 7, and Model 7 in Table 8).

Below, the models in which only the "adjusted income per capita" is included in the analyses, will be summarized. In the two region model, the odds ratio of the "adjusted income per capita" (INCOME_pc_) changed from 0.259 to 0.825 (which is much closer to the value of 1.0) (see Model 7 in Table 7). In other words, the change in the adjusted income per capita has much lower effect (almost nill) on the model in distinguishing provinces in the East versus in the West—than when it was not adjusted according to average household size (see Model 6 in Table 7).

Likewise, in the three region model, the odds ratio changed for the West from 5.444 to 1.523; and for the Central T. from 3.763 to 1.121 (which is very close to the value of 1.0) (see Model 6 in Table 8). In other words, the effect of the change in the adjusted income per capita decreased significantly; and it has almost no effect on the odds value for the Central T.

6. Conclusion

The aim of this study is to test the relative importance of the socio-demographic variables versus economic variables in the regional differentiation in Turkey. The univariate tests are repeated for all three types of regionalization of Turkey, i.e. in three (West, Central and Black Sea, which is called as Central T. for the sake of brevity, and East); and two regions (West and East). In the univariate analysis, East is the only region with values of socio-demographic variables less than the national average. However, in terms of the economic variables, Central T. as well as the East have lower values than the national average. Consequently, the inter-regional differences are attenuated especially in the economic variables in the two-region analysis where the Central and Black Sea regions are included in the so-called Western region.

Consistent results are obtained both from univariate and multivariate analyses where socio-demographic indicators especially those related to "fertility" and "education of females" were far more powerful than the economic indicators in distinguishing the regions from each other. Only "the average household size" could correctly classify as high as 93% of the provinces in the two-region model; and 81.7% in the three-region model. The optimal models were those which included only two variables: the average household size and the female literacy rate.

Among the socio-demographic variables, the health indicators of expected life at birth, and especially the infant mortality rates were found to be the least important. This indicates the ease of changing health parameters compared to the fertility and education of females which are highly affected by the particular cultural characteristics of the regions and can be changed in the relatively longer term.

The economic variables had the smallest coefficients in the principal component analysis; and they were not significant in the logistic regression tests.

Explorative, tentative analyses are carried out with the "income per capita" which proved to be even better than the average household size. The results should be viewed with due caution because of the quality of data. Subsequently, the income per capita was adjusted according to the case if the East (East and Southeast regions) had average

household size equal to the rest of the regions in Turkey i.e., if they had 5 instead of 7 members in their household. Such an adjustment resulted in 40% increase in the per capita incomes in the East. In the subsequent logistic regression analyses, the "adjusted income per capita" performed poorly; and the difference between regions, especially between the East and Central T. became nil. This finding indicates the significant effect of the past and present fertility rates on the economic variables. Further analyses together with the female literacy rates showed us that if the income per capita in the East was significantly improved, then the female literacy rate is the only dominant factor in distinguishing the regions in Turkey.

This study clearly showed that in the least developed region of Turkey, i.e. East and Southeast Turkey where huge amounts of public funds are being spent in terms of gigantic series of dams and irrigation projects, and other various infrastructure and services, equal effort and funds should also be spent for socio-demographic-cultural development of the region, especially in the areas of education of females and fertility reduction. Until the fertility rates are reduced to the national averages in these undeveloped regions, the efforts and expenditures to decrease the inter-regional disparity in per capita incomes, are futile.

ENDNOTES

(1) Southeast is the region where the GAP (Southeast Anatolian Project) have been applied which was initialized in 1977 and which is supposed to be finished in 2010. "The project has 22 dams, and 19 hydroelectric centers. The irrigation project targets 1.7 million hectares of land" (T.C.B.GAP B.K.I. Baskaligi, 2001, pp. 8-22).

(2) These two variables are most effective and statistically significant variables also in the subsequent analyses of logistic regression.

(3) Many values were about 100% which was because of in icluding 5 year olds in the numerator, but not in the denominator.

(4) The household income is obtained from the 1994 the household income distribution survey, and it is in terms of regional averages
(DIE 1999).

REFERENCES

Barro, R. J. (1991), Economic Growth in a Cross-Section of Countries, *Quarterly Journal of Economics*, 106: 407-43, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Caliskan, G., Ozelci,T., Varol,C. (1997), Turkiye'de Kentlesme Surecinin Bolgesel Farklilasmasi Analizi, *Planlama*, TMMOB Sehir Plancilari Odasi, 15: 22-32, Ankara.

Dellaportas, G. (1983), Classification of Nations as Developed and Less Developed: An Arrangement by Discriminant Analysis of Socioeconomic Data, *American Journal of Economics and Sociology*, 12/2: 153-66.

DIE (1999), *1994 Hanehalki Gelir Dagilimi Anketi Sonuclari, Sosyo-ekonomik Gostergeler, Turkiye Kent-kir ve Bolgeler.* DIE: Ankara.

DIE (06.07.2001), *Veri Sozlugu (Tanimlar)*, [http://www.die. gov.tr /TURKISH/ _SOZLUK/ data. html]

DIE (06.07.2001), *Temel Ekonomik ve Sosyal Gostergeler 1997-1998*, [http://www.die.gov.tr/ TURKISH/ _ISTATIS/Esg2/f.htm]

Dincer, B. (1996), *Ilcelerin Sosyo-ekonomik Gelismislik Siralamasi*, DPT: 4123,Ankara.

DPT (1998), *Illerin Sosyo-ekonomik Gelismislik Siralamasi Arastirmasi, 1996*, DPT: Ankara.

Garson, D., NC State University, PA765, Quantitative Research In Public Administration, [http://www2.chass.ncsu.edu/garson/pa765/discrim.htm]

Gedik,A., *CP524 Lecture Notes*, METU, 2001.

Ger, G. (1991), Utility of Social Indicators and Implications for Turkey, *Journal of Economics and Administrative Studies*, 5/1: 1-24.

Hirsch, F. (1977), *Social Limits to Growth*, London: Routledge & Kegan Paul, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Klecka, W.R. et al. (1975), *SPSS Primer*, NY: Mc Graw Hill Comp.

Kuznets, S. (1966), *Modern Economic Growth: Rate, Structure, and Spread*, New Haven, Conn.: Yale University Press, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Romer, P. M. (1989), Human Capital and Growth: Theory and Evidence, Working Paper 3173. Cambridge, Mass.: National Bureau of Economic Research, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Mrydal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Duckworth", in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Norusis, M.J. (1994), *SPSS Professional Statistics 6.1*

Singer, H. W. (1964), International Aid for Economic Development: Problems ad Tendencies, *International Development Review*, 6: 16-21, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Shaoguang, W., Angang, H. (1999), *Political Economy of Uneven Development: The Case of China*, NY: M.E.Starpe, Inc.

T.C.B.GAP B.K.I. Baskanligi (2001), *GAP 2001 Yili Faaliyet Raporu (GAP Annual Activity Report)*, Ankara.

UNDP (1998), *Human Development Report, Turkey*, Ankara. Yener, S., Koc, I. (2001), Turkiye'de Bolgeler Itibariyle Yasam Kalitesinin Olculmesi, in Nufus ve Kalkinma, 121-51, Ankara: Hacettepe Universitesi.

Young, A. (1991), Learning by Doing and the Dynamic Effects of International Trade, *Quarterly Journal of Economics*, 106/2: 369-406, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.

Young, A. (1993), Invention and Bounded Learning by Doing, *Journal of Political Economy*, 101/3: 443-72, in Rodriguez-Pose, A. (1998), *The Dynamics of Regional Growth in Europe: Social and Political Factors*, Oxford: Clarendon Press.