RUHR
ECONOMIC PAPERS

Rui Dang

A Decomposition Analysis of Cigarette Consumption Differences between Male Turkish Immigrants and Germans in West Germany 2002-2012

#602

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Jochen Kluve
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #602

Responsible Editor: Jochen Kluve

All rights reserved. Bochum, Dortmund, Duisburg, Essen, Germany, 2016

ISSN 1864-4872 (online) – ISBN 978-3-86788-699-4

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #602

Rui Dang

A Decomposition Analysis of Cigarette Consumption Differences between Male Turkish Immigrants and Germans in West Germany 2002-2012

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788699
ISSN 1864-4872 (online)
ISBN 978-3-86788-699-4

Rui Dang[1]

A Decomposition Analysis of Cigarette Consumption Differences between Male Turkish Immigrants and Germans in West Germany 2002–2012

Abstract

In this article, we investigate the differences in smoking behavior between male Turkish immigrants and male Germans, using data from the German Socio-Economic Panel (SOEP). More specifically, we use a Blinder-Oaxaca decomposition method for count data models, and isolate differences in the number of cigarettes consumed daily between Turkish immigrants and Germans into a component reflecting differences in observed socio-economic characteristics and a component reflecting unobserved smoking behavior. Our results reveal that more than 50% of the differences in cigarette consumption between male Turkish immigrants and male Germans is attributable to observable characteristics.

JEL Classification: J15, I14, C21

Keywords: Blinder-oaxaca decomposition; count data models; Turkish immigrants; smoking behavior differentials

January 2016

1 Rui Dang, RGS Econ and RWI. *– The author is grateful to Thomas K.Bauer, Deborah Cobb-Clark, Martin Siegel and Wang Wenjing for their valuable comments. The author also acknowledges the host of RWI and the financial support from the Ruhr Graduate School in Economics and the Mercator Foundation.* *- All correspondence to: Rui Dang, RGS Econ, Hohenzollernstr. 1-3, 45128, Essen, Germany, e-mail: rui.dang@rgs-econ.de*

1 Introduction

Smoking prevalence is higher in both Germany and Turkey than the OECD average of 20.7% (see OECD (2013)). According to the World Health Organization(WHO) and the Turkish Ministry of Health(MoH), the smoking prevalence is higher among Turkish males than their German counterparts, whereas the opposite applies to women.[1] As Turkish immigrants now form the largest ethnic minority in Germany[2], while often have a lower socioeconomic status than the host population and face segregation in education and the labor market (see, among others, Glitz (2014), Euwals et al. (2007) and Humpert (2014)), the disparity in cigarette consumption between Turkish immigrants and Germans have special importance. First, smoking is a major risk factor for malignancies, cardiovascular (e.g. atherosclerosis or myocardial infarction), respiratory diseases and cancers (see, among others, Vineis et al. (2004) and Doll et al. (2004)), in consequence, high smoking rates prompt high health care costs.[3] Second, the smoking disparity between male Turkish immigrants and male Germans may involve the phenomena that low-income households spend more on purchasing a product that is not good for health (see Pampel et al. (2010)).

Economic analysis explaining disparities in smoking mainly focus on differences in socioeconomic status. Previous research has shown that disparities in socioeconomic status are correlated to disparities in smoking behavior (see Cawley and Ruhm (2011) for a comprehensive review of relevant literature). Cutler and Lleras-Muney (2010) show that better educated people are less likely to smoke in the U.S., controlling for age, gender and parental background. Using the first wave of the German Health Interview and Examination Survey for Adults (DEGS), Lampert et al. (2013) also find differences in smoking behavior by social status (indexed by information on education and vocational training, occupational status and net equivalent income into a classification of low, middle and high status groups) for Germans aged 18-79, whereby women and men with low social status smoke roughly twice as often as those with a

[1] World Health Organization (2013) show that 30.5% of adult men in Germany smoke cigarettes and 26.4% are daily smokers, while 21.2% of female adult Germans are cigarette smokers and 17.6% are daily smokers. In Turkey, 41.3% of adult males are cigarette smokers and 37.3% are daily cigarette smokers, whereas only 13.0% of adult women are cigarette smokers and 10.7% are daily cigarette smokers (see Turkish Ministry of Health (2014)).

[2] In 2011, according to the national census, there were 2,956,000 people with current or previous Turkish nationality living in Germany, accounting for 18.5% of the German population with an immigration background and 3.6 % of the entire population in Germany (see Federal Offices for Migration and Refugees of Germany (2013)) the majority live in West Germany, including Berlin. Federal Statistical Office of Germany (2014) show that 2,771,000 Turkish immigrants lived in West Germany and Berlin in 2013, while only 22,000 Turkish immigrants lived in East Germany.

[3] Neubauer et al. (2006) estimated that the direct medical costs and total cost per smoker in Germany amounted to 346 Euros and 974 Euros in 2003, respectively, while the total cost of smoking amounted to 21 billion Euros per year for the German economy, including the indirect costs of occupational disability, early retirement and deaths related to smoking.

high social status, while they are also less likely to quit smoking.

In this paper, we focus on analyzing the smoking behavior of male Turkish immigrants and male Germans. The aims of our study are to (1) estimate the differences on average cigarette consumption per day among male Turkish immigrants and male Germans in Germany and (2) analyze the difference on average daily cigarette consumption by decomposing the differences in a part attributable to Turkish/German differences in observable characteristics and a part due to unobserved factors, i.e. unobserved smoking behavior. From a policy perspective, it is interesting to investigate whether the male Turkish/German differences in smoking behavior could be mainly due to observable characteristics or whether they are mainly explained by a different smoking behavior between those two population sub-groups. This knowledge could enable policy-makers to design anti-smoking policies more effectively. If differences on average daily cigarette smoking between male Turkish immigrants and male Germans could mainly be explained by observed differences in age structure, years of schooling, labor market and occupational status, number of children in the household and household income, policy-makers may not need to address specific target groups when designing tobacco control policies. By contrast, if differences in cigarette consumption are mainly due to unobserved smoking behavior, policy-makers may address specific target groups. The smoking behavior of Turkish immigrants might also be influenced by both the perceived social acceptance of smoking in Turkey and the adaptation towards the smoking patterns in Germany. The anti-smoking policies and those concerning integration in Germany may also influence a possible adaptation of Turkish immigrants.[4] Existing studies in terms of smoking among Turkish immigrants in Germany have found that smoking prevalence among Turkish immigrants converges to that of Germans with an increasing duration of stay (see, among others, Reeske et al. (2009) and Reiss et al. (2014))

This paper starts by using the Blinder-Oaxaca type decomposition proposed by Bauer et al. (2007) for mean differences of count data outcomes between two groups. Our empirical results reveal the existence of smoking behavior differentials between male Turkish immigrants and male Germans, whereby male Turkish immigrants smoked on average more than Germans.[5] Overall, our estimates indicate that more than 80% of mean daily cigarette consumption differentials among males can be explained by differences in observable characteristics. Conditional on being a smoker,the observable characteristics completely accounts for differences in number of cigarettes smoked per day. This results is robust across different regression models and model specifications.

[4]Though previous research using the German Socio-Economic Panel Study does not find that the introduction of smoke-free legislation in 2007 and 2008 change average smoking behavior within the German population, see e.g. Anger et al. (2011).

[5]There exists no significant difference in smoking prevalence between fmale Turkish immigrants and fmale Germans from our data

The remainder of this paper is structured as follows. In section 2, we present the data used for our empirical analysis, before we discuss the methodological approach in section 3. In section 4, we present the estimation results, section 5 concludes.

2 Data and Descriptive Statistics

Our empirical analysis employs data from the German Socio-Economic Panel (SOEP), which is a longitudinal survey of approximately 20,000 persons in 11,000 private households in the Federal Republic of Germany. The SOEP collects information on individuals' demographics, socioeconomic status, smoking behavior, income, etc. The survey includes weights to make the sample representative, given that it oversamples immigrants and high-income households. Further details can be found in Wagner et al. (2007).

The dependent variable is the average number of cigarettes consumed per day in the week before the interview, which is a count variable. For our main analysis, we focus exclusively on individuals who have valid information on the number of cigarettes smoked daily.[6] Smoking-related questions have been asked in the SOEP questionnaire in 1998, 1999, 2001 and every two years since 2002, although the average number of cigarettes consumed per day is not asked in the questionnaire for 1999. Furthermore, the consumption of cigarettes, pipes and cigars is not distinguished in the SOEP questionnaire for 2011. Accordingly, we make use of the SOEP data in 2002, 2004, 2006, 2008, 2010 and 2012. The identifier of Turkish immigrants is constructed from the following question in the SOEP: "What is your country of origin?". We treat individuals who answer "2" (Turkey as country of origin) as Turkish immigrants and those who answer "1" (Germany as country of origin) as Germans, exclusive of second-generation Turkish immigrants born in Germany.[7] We also drop all observations in East Germany because the SOEP does not include Turkish immigrants who live in this region. We pool data from six different waves, whereby our final sample comprises 1,216 person-year observations for male Turkish immigrants, and 36,031 for male German native-borns.[8]

[6] In our analysis, we only consider cigarette consumption, as two of the three outcome measures are quantity-based and no objective scale exists by which units of one type of tobacco could be sensibly converted into those of another. Regardless, cigar/pipe smoking is very rare in our data (only about 1% of individuals consume such tobacco products).

[7] Second generation Turkish immigrants include (i) persons who have been born in Germany but do not have German nationality; (ii) persons who have been born in Germany with German nationality whose parents have a foreign nationality or are both migrants; and (iii) persons who migrated to Germany before the age of 6

[8] The variables used in this paper were extracted using the add-on package PanelWhiz for Stata, written by Prof. John P. Haisken-DeNew (john@panelwhiz.eu). The PanelWhiz generated the do file to retrieve the SOEP data used in this research. Any data or computational errors are my own. See

Summary statistics by immigration status pooled for person-waves are presented in Table 1. The average number of cigarettes smoked per day in our sample is 8.997 with a variance of 11.081 for male Turkish immigrants. Male Germans on average consume 5.351 per day with a variance of 9.936. The raw data of our independent variable is thus overdispersed. 48.1 % percent of the male Turkish immigrants in our sample are regular smokers, while 29.3 % of male German native-borns are regular smokers. Throughout our empirical analysis, we control for age, years of schooling, indicator variables for being married, having at least one child. and net annual household income below 35,000 Euros, as well as four dummy variables indicating the labor market status (full-time job, part-time job, retired, and in training, with not participating labor market including registered unemployment acting as a reference group).

Table 1 further shows the differences in socioeconomic characteristics between male Turkish immigrants and male German natives. Significant differences in the means appear for many characteristics for which we control. Male Turkish immigrants in the sample are on average 3.645 years younger compared to German native-borns, have 2.421 years less schooling, around 9,132 Euros less household income, are more commonly married and live in urban regions or regions under urbanisation, more commenly have at least one child, less commonly have a full- or part-time job and are more commonly unemployed or inactive in the labor market. Most of these differences in observable characteristics are statistically significant across immigrant status. Table 2 summarizes the dependent variable. There are few large counts, with 71.2% of the sample taking the value of 0.

3 Empirical methods

We aim to isolate the part of the number of cigarettes smoked per day between male Turkish immigrants and male German native-borns that can be attributed to differences in observable characteristics, as well as the part explained by differences in estimated coefficients. The unexplained part will be interpreted as the component reflecting male Turkish immigrant/German differences in unobserved smoking behavior. Our outcome variable is the average number of cigarettes consumed per day, which takes the form of counts and is not normally distributed, as shown in Table 2. Such a violation of the assumption of normal distribution may yield either an overestimation or underestimation of the decomposition of the mean difference of daily cigarette consumption between male Turkish immigrants and Germans and its statistical significance (see, among others, Winkelmann (2000) and Cameron et al. (2013)). In order to obtain consistent parameter estimates, we employ the Blinder-Oaxaca type of decom-

Haisken-Denew and Hahn (2010) for more details about PanelWhiz.

position for count data models derived in Bauer and Sinning (2008) and Bauer et al. (2007)(see Sinning et al. (2008) for a detail description of the stata command nldecompose, which implements a Blinder-Oaxaca type of decomposition methods for non-linear models derived in Bauer et al. (2007) and Bauer and Sinning (2008)).

Consider the linear regression equation

$$CIG_{iG} = X_{iG}\beta_g + \epsilon_{iG} \tag{1}$$

where CIG_{iG} represents the number of cigarettes smoked daily by individual i in group G and X_{iG} is a vector of the observable characteristics described in Table 1. β_G denotes the vector of parameters and ϵ_{iG} is the error term. For the linear model (1), the decomposition proposed in Blinder (1973) and Oaxaca (1973) is:

$$\overline{CIG}_t - \overline{CIG}_g = \left[E_{\beta_t}\left(CIG_{it}|X_{it}\right) - E_{\beta_t}\left(CIG_{ig}|X_{ig}\right)\right] + \left[E_{\beta_t}\left(CIG_{ig}|X_{ig}\right) - E_{\beta_g}\left(CIG_{ig}|X_{ig}\right)\right] \tag{2}$$

where $\overline{CIG}_G = N_G^{-1}\sum_{i=1}^{N_G} C_{iG}$ and $\overline{X}_G = N_G^{-1}\sum_{i=1}^{N_G} X_{iG}$ for G=t,g. Accordingly, $\overline{CIG}_t$ is the average number of cigarettes smoked per day for the male Turkish immigrants and $\overline{CIG}_g$ is the average number of cigarettes smoked per day for German native-borns, $\overline{CIG}_t - \overline{CIG}_g$ is defined as the mean difference in the number of cigarettes smoked daily, and $\left[E_{\beta_G}\left(CIG_{iG}|X_{iG}\right)\right]$ refers to the conditional expectation of CIG_{iG}. The first term of equation (2) $E_{\beta_t}\left(CIG_{it}|X_{it}\right) - E_{\beta_t}\left(CIG_{ig}|X_{ig}\right)$ represents the proportion of the mean difference in the number of cigarettes smoked per day that can be explained by differences in each observable characteristics, while the remaining components on the right hand side of equation (2) $E_{\beta_t}\left(CIG_{ig}|X_{ig}\right) - E_{\beta_g}\left(CIG_{ig}|X_{ig}\right)$ represents the proportion of the mean BMI gap that cannot be explained by observable characteristics.[9]

For non-negative count outcomes $\overline{CIG}_G$ in equation (2), a model with Poisson distribution or Negative binomial distribution is much more appropriate than an ordinary least-squares linear model (see, among others,Winkelmann (2000) and Cameron et al. (2013)). Poisson and negative binomial (hereafter Negbin) regressions are often used to model count data, given that the independent variable departs from the Poisson distribution due to over-dispersion, i.e. the conditional variance exceeds the conditional mean, as shown in Table 1. Negbin regression can be an alternative to the Poisson regression, since Negbin distribution has an extra parameter to model the over-dispersion, in addition to having the same mean structure as a Poisson regression.

Table 2 also indicates that the distribution of the number of cigarettes smoked per day has a much larger than expected number of observed zeros than assumed by Pois-

[9]Bauer et al. (2007) points out that in a linear regression model, equation (2) is equivalent to standard Blinder-Oaxaca decomposition formula in the linear setting: $\overline{CIG}_t - \overline{CIG}_g$=$\left(\overline{X}_t - \overline{X}_g\right).\hat{\beta}_t + \overline{X}_t(\hat{\beta}_t - \hat{\beta}_g)$.

son and Negbin distribution. Many smokers may not smoke at all or only rarely, e.g. once a week. Zero-inflated and hurdle models (each assuming either the Poisson or negative binomial distribution of the outcome) have been developed to cope with zero-inflated outcome data with (negative binomial) or without (Poisson distribution) over-dispersion. Both zero-inflated and hurdle models deal with the high occurrence of zeros in the observed data, although there is one important distinction in terms of how they interpret and analyze zero counts.

Therefore, we also apply zero-inflated [10] and hurdle models[11], in addition to the Poisson and Negbin models. Both zero-inflated and hurdle models assume either the Poisson or negative binomial distribution of the outcome and attempt to account for excess zeros. The distinction between zero-inflated and hurdle models is that there is a single type of zeros in the latter, whereas there are two types of zeros in the former: zero outcomes arise from Poisson or negative binomial distributions that include both zero and non-zero counts (the "sampling zeros" or "true zeros") and from zero counts (the "structural zeros" or "true zeros"). Zero-inflated and hurdle models can yield different results with very different interpretations. We need to decide which is more appropriate for the nature of distribution in our data. In our empirical analysis, a zero-inflated model assumes that some individuals smoke zero cigarettes per day because they are non-smokers, while others individuals smoke but score zero because their smoking behavior is assumed to be on a Poisson or Negbin distribution that includes both zero and non-zero counts of cigarettes. Zero-inflated models estimate one (Poisson or Negbin) equation for the count model and another for the excess zeros. In contrast to zero-inflated models, a hurdle model combines a dichotomous model for the binary outcome of the count being below or above the hurdle (most widely set at 0), with a truncated model for outcomes above the hurdle.

To apply a Blinder-Oaxaca decomposition method to different count data models described above, we make use of the respective sample counterparts $S(\hat{\beta}_G, X_{iG})$ of the conditional expectation of CIG_{iG}, i.e.

$$E_{\beta_G}(CIG_{iG}|X_{iG}) = S(\hat{\beta}_G, X_{iG}), G = t, g$$

which are defined in Bauer et al. (2007) and Bauer and Sinning (2008).

Hence, the decomposition equation (2) can be written as

$$\overline{CIG}_t - \overline{CIG}_g = \left[S(\hat{\beta}_t, X_{it}) - S(\hat{\beta}_g, X_{ig})\right] + \left[S(\hat{\beta}_t, X_{ig}) - S(\hat{\beta}_g, X_{ig})\right], G = t, g \quad (3)$$

[10] Zero-inflated models are firstly discussed in Lambert (1992)(see also Winkelmann (2000) and Cameron et al. (2013) for more details and empirical applications).

[11] Hurdle models are firstly discussed in Mullahy (1986) (see also (see also Winkelmann (2000) and Cameron et al. (2013) for more details and empirical applications).

Bauer et al. (2007) provide an overview of sample counterparts in non-linear models, as shown in Table 3.

4 Decomposition Results

We estimate the five count data models described in section 3 separately for both male Turkish immigrants and male Germans, i.e. we estimate Poisson and Negbin models as well as zero-inflated Poisson and Negbin models, plus the two part hurdle Negbin model. We use likelihood-ratio tests and Vuong tests (Vuong (1989)) for non-nested models and Akaike information criterion (AIC) to test the different models against each other. The descriptive statistics reported in Tables 1 and 2 already imply that there is overdispersion in the dependent variable. Therefore, all of our tests reject the different Poisson model and zero-inflated Poisson in favor of the Negbin-models for both male Turkish immigrants and male Germans. Testing the Negbin models and Hurdle Negbin models against the zero-inflated Negbin model using the Vuong test (see Vuong (1989)) finally shows that the zero-inflated Negbin model describes the data best for both male Turkish immigrants and male Germans. The zero-inflated Negbin model assumes that individuals reporting zero consumption of cigarettes may be either non-smokers who never smoke, or potential smokers who quit smoking temporarily.

Table 4 presents the estimation results of the zero-inflated Negbin model for both male Turkish immigrants and male Germans. A number of factors may contribute to the male Turkish immigrants' higher level of daily cigarette consumption. The reported standard errors are clustered at the person level.

For male Turkish immigrants and male Germans, the potential of being a non-smoker increases with age and years of education. Compared to male Germans who are not participating labor market, individuals in educational training and retirement have a higher probability of being a non-smoker, while male Germans with a full-time employment have a significantly lower probability of being a non-smoker. Married male Germans are significantly more likely to be non-smokers than non-married men. Male Germans with an annual household income below 35,000 Euros have a significantly higher probability of being smoker than those with an annual houshehold income above 35,000 Euros. There are some remarkable differences between male Germans and their Turkish immigrant counterparts concerning the probability of being a non-smoker, although the respesctive coefficients for male Turkish immigrants are statistically insignificant. Only age and years of schooling reduces the probability of being a non-smoker for male Turkish immigrants.

Conditional upon being a potential smoker, the number of cigarettes smoked per day increases with age and but decreases with years of education for male Germans, while male German workers with full-time employment smoke significantly more cigarettes than those not participating in the labor market, while Male Germans in educational training and retirement smoke significantly less than those not in the labor market. Males Germans with an annual household income below 35,000 Euros smoke more cigarettes than those with an annual household income above 35,000 Euros. For male Turkish immigrants, the coefficients concerning age, years of schooling, marital status, household income and indicator variable for having a children in the household are not statistically significant, while only having full-time employment increases the number of cigarettes smoked daily.

Table 5 shows that the differences in the daily cigarette consumption between male Turkish immigrants and male Germans are mostly due to differences in observable characteristics rather than differences in coefficients. Referring to the zero-inflated Negbin model - the model that best describes our data best - 54.5% of the difference could be explained by differences in observable characteristics and 45.5 % by differences in coefficients. The decomposition analyses are rather stable across the four different models.

Table 6 reports the estimation results of the hurdle model. The decomposition carried out in the hurdle model includes a logit specification that models the binary decision to smoke vs not smoking and a truncated Negbin model of average daily cigarette consumption for the subsample of smokers[12]. According to Table 6, 49.2% of the differential concerning the decision to smoke or not between male Turkish immigrants and male Germans is due to differences in observable characteristics, while about half of the differential (50.8%) is due to differences in the coefficients. Among smokers, 92.5% of the differential concerning the decision to smoke or not between male Turkish immigrants and male Germans is due to differences in observable characteristics, while 7.5% of the differential is due to differences in the unobservable.

5 Conclusion

Given that male Turkish immigrants in Germany smoke more on average than German natives, this paper provides a detailed analysis of the determinants of cigarette consumption among male Turkish immigrants and Germans. We employ the Blinder-Oaxaca type decomposition method for count data developed by Bauer et al. (2007) to decompose the mean difference in the number of cigarettes smoked daily between

[12]The point estimates and the significance levels of the estimated coefficients of the logit and truncated Negbin models are quite close to those reported in Table 4, Hence, Table 6 only reports the results of the decomposition analysis.

these two population sub-groups.

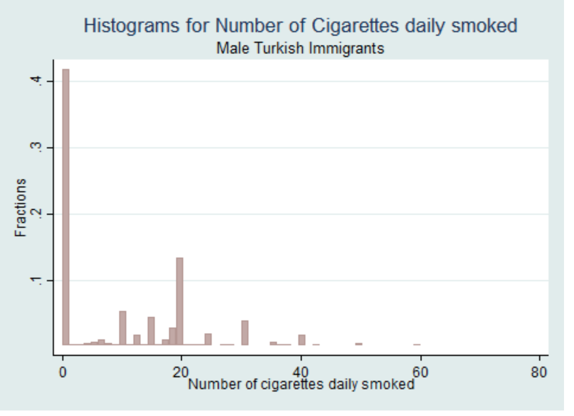
Our empirical results confirm that adult male Turkish immigrants consume more cigarettes on average than native Germans in Germany. Our empirical results further show that about 50% of this difference can be explained by observable characteristics. Concerning the decision of whether to smoke or not, 49.2% of the male Turkish immigrant/ German differentials in the probability of being a smoker can be explained by differences in observable characteristics, whereas the remaining 50.8% are left unexplained. Conditional upon being a smoker, the differences in observable characteristics explain 92.5 % of the Turkish immigrant/German differences in the number of cigarettes consumed per day. Therefore, among smokers, the explaining portion is the element that accounts for most of differences in the number of cigarettes smoked per day between male Turkish immigrants and male Germans.

The main policy implications of our results are as follows. Anti-smoking policies may need to address male Turkish immigrants and male Germans differently, given that aroud 50% of the differences in the number of cigarettes smoked per day at the mean are attributable to differences in unobserved smoking behavior. These policy measures may be designed more effectively if behavioral differences would be taken into account. Policy-makers may also need to distinguish between anti-smoking campaigns aimed at quitting smoke and those intended to reduce cigarette consumption among smokers. Our results suggest that anti-smoking policies to reduce smoking prevalence may need to address male Turkish immigrants and male Germans differently, while anti-smoking policies to reduce the conditional demand for cigarettes may not need to address male Turkish immigrants and male Germans differently.

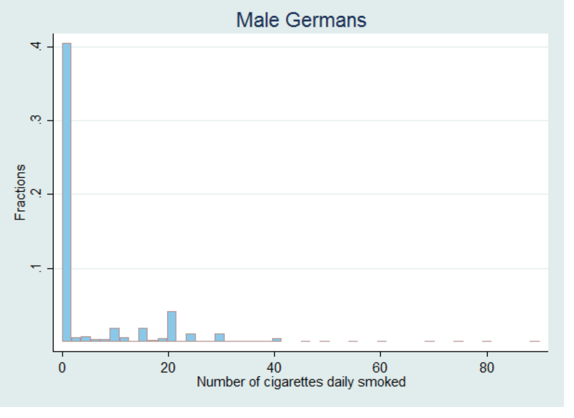
Table 1: Descriptive Statistics

	Turkish Immigrants (1)	German native-borns (2)	Differences (1)-(2)
Smoke	0.481	0.293	0.188***
	(0.5)	(0.455)	(0.020)
Number of cigarettes	8.997	5.351	3.645***
	(11.081)	(9.936)	(0.426)
Age	46.192	50.682	-4.490***
	(13.737)	(17.255)	(0.557)
Years of schooling	9.802	12.223	-2.421***
	(2.003)	(2.67)	(0.087)
Married (dummy)	0.865	0.587	0.278***
	(0.342)	(0.492)	(0.014)
Children (dummy)	0.603	0.238	0.365***
	(0.49)	(0.426)	(0.013)
Full time job (dummy)	0.535	0.561	-0.026
	(0.499)	(0.496)	(0.015)
Part time job (dummy)	0.030	0.051	-0.021***
	(0.17)	(0.22)	(0.006)
Training (dummy)	0.010	0.022	-0.012**
	(0.099)	(0.146)	(0.004)
Registered unemployed (dummy)	0.174	0.048	0.125***
	(0.379)	(0.215)	(0.007)
Not participating in labor market (dummy)	0.044	0.031	0.014***
	(0.206)	(0.172)	(0.005)
Retired (dummy)	0.204	0.261	-0.0579***
	(0.403)	(0.439)	(0.0013)
Annual Household Income(1,000 Euro)	28.566	37.699	-9.132***
	(15.597)	(34.909)	(0.604)
Log(Annual Household Income)	10.119	10.35	-0.230***
	(0.58)	(0.647)	(0.023)
Annual Houshel Income less than 35,000 Euro	0.721	0.554	0.168***
	(0.449)	(0.497)	(0.019)
Annual Houshel Income 35,000-54,000 Euro	0.234	0.285	-0.053***
	(0.424)	(0.452)	(0.018)
Annual Houshel Income more than 54,000 Euro	0.044	0.160	-0.115***
	(0.206)	(0.367)	(0.008)
Urban or under urbanization (dummy)	0.966	0.836	-0.130***
	(0.181)	(0.37)	(0.011)
Rural (dummy)	0.034	0.164	-0.130***
	(0.181)	(0.37)	(0.108)
N	1216	36031	

Means and standard deviations are weighted using the SOEP weight. Standard deviations are reported in parentheses, the significance level of the mean differences are calculated using a t-test. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.



(a) Male Turkish Immigrants



(b) Male Germans

Figure 1: Number of Cigarettes Daily Smoked

NOTE.— Figure 1 illustrates the histogram of number of cigarettes daily smoked for both Male Turkish immigrants and Male German native-borns. Figure 1(a) shows for male Turkish immigrants and 1(b) shows for male Germans. Figures 1(a) and 1(b) clearly indicates that . Data source: German Socio-Economic Panel(GSOEP)..

Table 2: Number of cigarettes daily smoked

Count	0	1	2	3	4	5	6	7	8	9	≥ 10	Total
Male Turkish Immigrants												
Frequency	608	2	4	5	6	9	8	7	6	3	558	1,216
Relative frequency	0.50	0.002	0.003	0.004	0.005	0.007	0.007	0.006	0.005	0.003	0.459	1
Male Germans												
Frequency	26,076	153	200	183	107	407	142	98	244	38	8383	36,031
Relative frequency	0.724	0.004	0.006	0.005	0.003	0.011	0.004	0.003	0.007	0.001	0.319	1

Table 3: Sample Counterparts for count data models

Model	Sample counterparts
Logit	$1S(\hat{\beta}_G^{Logit}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} \wedge(X_{iG}\hat{\beta}_G)$, where $\wedge$ is the cumulative logistic density, function
Poisson	$S(\hat{\beta}_G^P, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} exp(X_{iG}\hat{\beta}_G^P)$
Hurdle Poisson	$S(\hat{\beta}_G^{HP}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} \frac{exp(\hat{\beta}_G^{HP} X_{iG})}{(1-exp(-exp(\hat{\beta}_G^{HP} X_{iG})))(1+exp(\hat{\Gamma}_G Z_{iG}))}$
Zero-inflated Poisson	$S(\hat{\beta}_G^{ZIP}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} \frac{exp(\hat{\beta}_G^{ZIP} X_{iG}}{1+exp(\hat{\Gamma} Z_{iG})}$
Negbin	$S(\hat{\beta}_G^{NB}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} exp(X_{iG}\hat{\beta}_G^{NB})$
Hurdle Negbin	$S(\hat{\beta}_G^{HNB}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} \frac{exp(\hat{\beta}_G^{HNB} X_{iG})}{(1-(1+\alpha exp(-exp(\hat{\beta}_G^{HNB} X_{iG}))^{-\frac{1}{\alpha}}))(1+exp(\hat{\Gamma}_G Z_{iG}))}$
Zero-inflated Negbin	$S(\hat{\beta}_G^{ZINB}, X_{iG}) = \frac{1}{N_G}\sum_{i=1}^{N_G} \frac{exp(\hat{\beta}_G^{ZINB} X_{iG}}{1+exp(\hat{\Gamma} Z_{iG})}$

Table 4: Zero-inflated Negbin Esitmates of determinants of the No. cigarettes daily smoked (SOEP 2002-2012)

	Male Turkish Immigrants		Male Germans	
	Logit model(inflate)	Truncated Negbin	Logit model(inflate)	Truncated Negbin
	Probability on Non-smokers	No. of Cigarettes daily smoked	Probability on Non-smokers	No. of Cigarettes daily smoked
Age	0.0359**	0.003	0.017***	0.008***
	(0.0126)	(0.004)	(0.002)	(0.001)
Years of education	0.0113*	-0.021	0.150***	-0.036 ***
	(0.0573)	(0.017)	(0.010)	(0.005)
Married (dummy)	0.391	-0.268	0.316***	-0.040
	(0.285)	(0.0717)	(0.055)	(0.021)
Children (dummy)	0.080	-0.041	-0.077	0.009
	(0.250)	(0.063)	(0.050)	(0.020)
Full time job (dummy)	0.338	0.114*	-0.148*	0.067
	(0.248)	(0.050)	(0.102)	(0.022)
Part time job (dummy)	-0.0776	-0.097	0.117	0.009
	(0.513)	(0.118)	(0.087)	(0.039)
In educational training (dummy)	1.314	-0.042	0.445**	-0.124 **
	(0.762)	(0.135)	(0.095)	(0.038)
Retired (dummy)	0.404	-0.109	0.179**	0.214***
	(0.381)	(0.109)	(0.091)	(0.035)
Annual HH. Income less than 35,000 Euro (dummy)	-0.162	0.018	-0.347***	-0.052***
	(0.194)	(0.110)	(0.055)	(0.016)
Rural regions (dummy)	-0.768	-0.768	0.042	-0.045**
	(0.526)	(0.526)	(0.060)	(0.024)
Constant	-3.513***	-3.513***	-1.1891***	2.991***
	(0.977)	(0.977)	(0.150)	(0.067)
α	-1.792***		-1.364***	
	(0.114)		(0.035)	
Vuong: ZINB vs Standard Negbin and Hurdle Negbin	20.39		65.84	
p-Value	0		0	
Wald-Statistic LR test(χ^2): ZINB vs ZIP	1016.31		24000	
Log pseudolikelihood	-2901.04		-55530.35	
No.Obs.	1216		36031	

Standard errors in parentheses are clustered at individual level. Reference group is a non-married individual, not participating in labor market with an annual HH. income more than 35,000 Euro. The regression further includes year dummies. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Decomposition Results

	Poison		Zero-inflated Poisson		Negbin		Zero-inflated Negbin	
	Coefficient	in % of $\widehat{\Delta}$	Coefficient	in % of $\widehat{\Delta}$	Coefficient	in % of $\widehat{\Delta}$	Coefficient	in % of $\widehat{\Delta}$
				Male				
$\widehat{\Delta}$	4.693***		4.689***		4.706***		4.685***	
	(0.344)		(0.359)		(0.317)		(0.353)	
Explained part	2.671***	56.9%	2.548***	54.3%	2.336***	49.8%	2.551***	54.5%
	(0.141)		(0.140)		(0.155)		(0.127)	
Unexplained part	2.022***	43.1%	2.140***	45.7%	2.350***	50.2%	2.134***	45.5%
	(0.364)		(0.349)		(0.331)		(0.332)	

Bootstrap standard error (100 replications)in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Decomposition results two-part model(SOEP 2002-2012)

	Male			
	Logit		Truncated Negbin	
	Coefficient	in % of $\widehat{\Delta}$	Coefficient	in % of $\widehat{\Delta}$
$\widehat{\Delta}$	0.224	100%	1.386	100%
Explained	0.110***	49.2%	1.285***	92.5%
	(.006)		(0.151)	
Unexplained	0.114***	50.8%	0.101	7.5%
	(0.016)		(0.437)	

Bootstrap standard error (100 replications)in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Anger, S., M. Kvasnicka, and T. Siedler (2011). One last puff? public smoking bans and smoking behavior. *Journal of Health Economics 30*(3), 591 – 601.

Bauer, T., S. Göhlmann, and M. Sinning (2007). Gender differences in smoking behavior. *Health Economics 16*(9), 895–909.

Bauer, T. and M. Sinning (2008, May). An extension of the Blinder−Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis 92*(2), 197–206.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources 8*(4), pp. 436–455.

Cameron, A. C., P. Hammond, A. Holly, and P. K. Trivedi (2013). *Regression Analysis of Count Data*. Cambridge: Cambridge Univ. Press.

Cawley, J. and C. J. Ruhm (2011). Chapter three - the economics of risky health behaviors. In T. G. M. Mark V. Pauly and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2 of *Handbook of Health Economics*, pp. 95 – 199. Elsevier.

Cutler, D. M. and A. Lleras-Muney (2010, January). Understanding differences in health behaviors by education. *Journal of Health Economics 29*(1), 1–28.

Doll, R., R. Peto, J. Boreham, and I. Sutherland (2004). Mortality in relation to smoking: 50 years' observations on male british doctors. *BMJ 328*(7455), 1519.

Euwals, R., J. Dagevos, M. Gijsberts, and H. Roodenburg (2007, March). The Labour Market Position of Turkish Immigrants in Germany and the Netherlands: Reason for Migration, Naturalisation and Language Proficiency. IZA Discussion Papers 2683, Institute for the Study of Labor (IZA).

Federal Offices for Migration and Refugees of Germany (2013). Migrationsbericht des bundesamtes für migration und flüchtlinge im auftrag der bundesregierung migrationsbericht. Technical report, Germany Federal Office of Migration and Refugees.

Federal Statistical Office of Germany (2014). Bevölkerung und erwerbstätigkeit bevölkerung mit migrationshintergrund - ergebnisse des mikrozensus 2013 - [population and occupation. people with migration background - results from the micro census 2013]. Technical report, Federal statistical office of Germany.

Glitz, A. (2014). Ethnic segregation in germany. *Labour Economics 29*(C), 28–40.

Haisken-Denew, J. and M. Hahn (2010). Panelwhiz: Efficient data extraction of complex panel data sets-an example using german soep. *Schmollers Jahrbuch 130*(4), 643 – 654.

Humpert, S. (2014). Trends in occupational segregation: What happened with women and foreigners in germany? *European Economics Letters 3*(2), 36–39.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), pp. 1–14.

Lampert, T., E. von der Lippe, and S. Mäters (2013). Verbreitung des rauchens in der erwachsenenbevölkerung in deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 56*(5-6), 802–808.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics 33*(3), 341 – 365.

Neubauer, S., R. Welte, A. Beiche, H.-H. Koenig, K. Buesch, and R. Leidl (2006). Mortality, morbidity and costs attributable to smoking in germany: update and a 10-year comparison. *Tobacco Control 15*(6), 464–471.

Oaxaca, R. (1973, October). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.

OECD (2013). Tobacco consumption among adults. In *Health at a Glance 2013*, pp. 54 – 55.

Pampel, F. C., P. M. Krueger, and J. T. Denney (2010, August). Socioeconomic Disparities in Health Behaviors. *Annual review of sociology 36*, 349–370.

Reeske, A., J. Spallek, and O. Razum (2009). Changes in smoking prevalence among first- and second-generation turkish migrants in germany - an analysis of the 2005 microcensus. *International Journal for Equity in Health 8*(1), 26.

Reiss, K., O. Sauzet, J. Breckenkamp, J. Spallek, and O. Razum (2014). How immigrants adapt their smoking behaviour: comparative analysis among turkish immigrants in germany and the netherlands. *BMC Public Health 14*(1).

Sinning, M., M. Hahn, and T. K. Bauer (2008, December). The Blinder - Oaxaca decomposition for nonlinear regression models. *Stata Journal 8*(4), 480–492.

Turkish Ministry of Health (2014). Global adult tobacco survey turkey 2012.

Vineis, P., M. Alavanja, P. Buffler, E. Fontham, S. Franceschi, Y. T. Gao, P. C. Gupta, A. Hackshaw, E. Matos, J. Samet, F. Sitas, J. Smith, L. Stayner, K. Straif, M. J. Thun, H. E. Wichmann, A. H. Wu, D. Zaridze, R. Peto, and R. Doll (2004). Tobacco and cancer: Recent epidemiological evidence. *Journal of the National Cancer Institute 96*(2), 99–106.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), pp. 307–333.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The german socio-economic panel study (soep): Scope, evolution and enhancements. SOEPpapers on Multidisciplinary Panel Data Research 1, DIW Berlin, The German Socio-Economic Panel (SOEP).

Winkelmann, R. (2000). *Econometric analysis of count data* (3rd ed.). Springer-Verlag, Berlin.

World Health Organization (2013). Who report on the global tobacco epidemic 2013 - enforcing bans on tobacco advertising, promotion and sponsorship.

Table 7: Variable Descriptions

Variable	Description
Number cigarettes	Number of cigarettes daily smoked
Age	Age of individual in years
Years of education	Years of individual's education
Full time job	1 if individual has a full time job including civil-/military service; 0 otherwise
Part time job	1 if individual has a part time job;0 otherwise
Training	1 if individual is in vocational training; 0 otherwise
Not in the Labor Market	1 if individual is currently registered unemployed, or individual who is neither in retirement nor registered unemployed and does not participate at the labor market; 0 otherwise
Retired	1 if individual is older than 65 or receives pension 0 otherwise
Married(dummy)	1 if individual is married ;0 otherwise
Children(dummy)	1 if individual has at least one child;0 otherwise
Annual HH income less than or equal to 35,000 Euro	1 if individual's annual HH. income is less than or equal to 35,000 Euro; 0 otherwise
Rural	1 if individual resides in a county catogorized as rural areas; 0 otherwise
Urban or Under Unbanization	1 if individual resides in a county catogorized as urban areas or areas under urbanisation; 0 otherwise