TAUB CENTER
FOR SOCIAL POLICY STUDIES IN ISRAEL

RESEARCH PAPER SERIES

IS THE ISRAELI LABOR MARKET SEGMENTED?

Ezra Fishman and Ayal Kimhi

Research Paper No. 13-01

* * *

Ezra Fishman is a Ph.D. student at the Population Studies Center, University of Pennsylvania. Ayal Kimhi is the Deputy Director of the Taub Center for Social Policy Studies in Israel, a Professor of Agricultural Economics and Management at the Hebrew University of Jerusalem, and the Director of The Center for Agricultural Economic Research. *All errors are the authors' own. The views expressed herein are those of the authors and do not necessarily reflect the views of the Taub Center for Social Policy Studies in Israel.*

The Taub Center was established in 1982 under the leadership and vision of Herbert M. Singer, Henry Taub, and the American Jewish Joint Distribution Committee, the Center is funded by a permanent endowment created by the Henry and Marilyn Taub Foundation, the Herbert M. and Nell Singer Foundation, Jane and John Colman, the Kolker-Saxon-Hallock Family Foundation, the Milton A. and Roslyn Z. Wolf Family Foundation, and the American Jewish Joint Distribution Committee. For more information go to www.taubcenter.org.il or write to us at info@taubcenter.org.il.

מרכז טאוב
לחקר המדיניות החברתית בישראל

סדרת ניירות מחקר

האם שוק העבודה בישראל מקוטע?

עזרא פישמן ואיל קמחי

נייר מחקר מס׳ 13-01

תקציר

אנו נוקטים שיטה של רגרסיה מעורבת על מנת לאפיין מקטעים נפרדים בשוק העבודה בישראל. במודל בעל שלושה מקטעים אנו מבחינים בין מקטע שבו השכר נמוך ומקטע שבו השכר גבוה, וכן מקטע נוסף בעל שונות שכר גבוהה שלא ניתן לכלול אותו באחד המקטעים האחרים. זה מאפשר לנו להשוות בין המקטעים של שכר נמוך ושכר גבוה לאחר "ניקוי" המדגם מתצפיות שמוסיפות בעיקר "רעש" סטטיסטי. סימולציה של המקטעים באמצעות ההסתברות החזויה המקסימלית של כל תצפית, כפי שנהוג בספרות, הביא למקטעים שונים משמעותית בגודלם מהגדלים שהתקבלו בתוצאות האמידה. אנו מציעים שיטה חלופית שבה כל התצפיות משוקללות בהסתברויות החזויות ונכללות באופן זה בכל אחד מן המקטעים. באמצעות שיטה זו אנו מוצאים הבדלים קטנים אך הגיוניים בין המאפיינים של העובדים בשני המקטעים האמורים. ההבדלים בין המקדמים של משוואות השכר, המהווים את התמורות למאפייני העובדים, היו גדולים הרבה יותר. לדוגמה, נשים יהודיות וגברים ערבים משתכרים פחות במקטע השוק שבו השכר גבוה, ואילו נשים ערביות משתכרות פחות במקטע שבו השכר נמוך. התמורה להשכלה גבוהה משמעותית ליהודים העובדים במקטע שבו השכר גבוה, אולם היא קיימת לנשים ערביות רק במקטע שבו השכר נמוך, ואינה קיימת כלל לגברים ערבים. התוצאות מעידות על כך שחלק הארי של פערי השכר בישראל נוצר על ידי גורמים שאינם נצפים ולא על ידי מאפיינים נצפים. התוצאות גם מובילות למספר תובנות הרלוונטיות למדיניות על הקשרים שבין השכלה, מיעוטים אתניים ושכר.

* * *

עזרא פישמן הוא דוקטורנט במרכז ללימודי אוכלוסייה, אוניברסיטת פנסילבניה. איל קמחי הוא פרופסור לכלכלה חקלאית באוניברסיטה העברית, סמנכ״ל מרכז טאוב לחקר המדיניות החברתית בישראל, ומנהל המרכז למחקר בכלכלה חקלאית. כל הטעויות הן של המחברים. הדעות המובאות להלן הן של המחברים ואינן בהכרח משקפות את דעות מרכז טאוב לחקר המדיניות החברתית בישראל.

המרכז נוסד ב-1982 ביוזמתם של הרברט מ' סונגר, הנרי טאוב, וארגון הג'וינט האמריקאי. המרכז ממומן באמצעות קרן צמיתה שהוקמה על-ידי קרן הנרי ומרילין טאוב, קרן הרברט ונל סינגר, ג'ין וג'ון קולמן, קרן משפחת קולקר-סקסון-הלוק, קרן משפחת מילטון א' ורוזלין ז' וולף, וארגון הג'וינט האמריקאי. למדע נוסף ראו taubcenter.org.il או בדוא"ל info@taubcenter.org.il.

Is the Israeli Labor Market Segmented?

Ezra Fishman and Ayal Kimhi

Abstract

We use a mixture regression model to identify different segments in the Israeli labor market. In the three-segment model, we obtained a low-wage segment and a high-wage segment, as well as a third segment with a large wage variability that cannot be assigned to either of the other two segments. This allows us, in effect, to compare the low-wage and high-wage segments after "purging" the "noisy" observations. Assigning workers to simulated segments using the maximum estimated posterior probability led to segment shares that are meaningfully different than the estimated population shares of each segment. We propose an alternative method in which observations are weighted by their posterior probabilities and then included in all simulated segments. Using this method, we found quantitatively small but qualitatively reasonable differences in the characteristics of workers between the low-wage segment and the high-wage segment. The between-segment differences in wage equation coefficients, representing the returns to these worker attributes, were much larger than the differences in worker attributes themselves. For example, Jewish females and Arab males suffer considerable wage penalties in the high-wage segment, while Arab females suffer wage penalties in the low-wage segment. Returns to schooling are considerably higher for Jewish workers in the high-wage segment, while they are positive for Arab females only in the low-wage segment, and do not exist for Arab males in either segment. Altogether, the results indicate that much of the wage disparities in Israel are due to unobserved factors rather than to observable characteristics. They also lead to some policy-relevant insights about the links between schooling, ethnic minorities and wages.

Ezra Fishman is a Ph.D. student at the Population Studies Center, University of Pennsylvania. Ayal Kimhi is the Deputy Director of the Taub Center for Social Policy Studies in Israel, a Professor of Agricultural Economics and Management at the Hebrew University of Jerusalem, and the Director of The Center for Agricultural Economic Research.

Introduction

The tent protests across Israel in the summer of 2011 highlighted a widespread perception that achieving a middle-class lifestyle in Israel is increasingly difficult. The protesters focused on rising socioeconomic inequality in Israel. An increase in inequality can affect the middle class in two main ways: it can reduce its size, moving people either up or down the income distribution, and it can increase the gap between the middle class and the upper class. It seems like the latter was the main concern of Israeli protesters.

The first question that comes to mind in this context is what exactly is the middle class? The simplest way to define the middle class is as a range (not necessarily symmetric) of incomes around the median (Birdsall et al., 2000). Other definitions include references to wealth and occupation (Atkinson and Brandolini, 2011). All these definitions are arbitrary, and choosing one definition over another may yield a different conclusion. Some methodologies enable analyses of changes in the middle class without explicitly defining it. These include polarization measures, which examine the between-group inequality compared to the within-group inequality in a population that is made of an arbitrary number of groups (Esteban and Ray, 1994), as well as examination of changes in entire distributions (Burkhauser et al, 1999).

For developing countries, it is common to treat all those above a certain income or consumption level as the middle class (Ravallion, 2010). This apparently assumes that the upper class is negligible or not interesting. In Israel, it is also quite common to relate to two classes only. Lubell (2011) mentions the concept of "two States of Israel" – one economically dynamic, at the forefront of the modern global economy, and a high standard of living; the other struggling to get by. Earlier, the concept of “first Israel” and “second Israel” was related to Jews of Ashkenazi (western) origin and those of Sefaradi (eastern) origin. Although Haberfeld and Cohen (2007) found that earnings differences between Ashkenazi men and other gender and ethnic groups have not narrowed since the 1970s, it is safe to say that divisions across origins are coming close to be a matter of the past in Israel. This paper explores the question, to what extent is the "two Israels" idea manifest in the labor market? Specifically, we use the idea of labor market segmentation to ask to what extent wage earners in Israel can be divided into the "haves" and the "have-nots." We will also allow for more than two labor market segments, without pre-specifying the number of segments, in order to examine the existence of a middle class.

In neoclassical labor economics, the labor market is treated as one single, heterogeneous market, where the differential productivity of workers explains differences in wages. That is, workers are paid in proportion to their productivity. In contrast, labor market segmentation theory suggests that in the simplest case, there are two labor markets: the primary labor market has good jobs, good working conditions, and high returns to human capital, but barriers to entry; the secondary labor market has bad jobs, bad working conditions, and low returns to human capital.

Much of the literature on the topic has focused on trying to find evidence of segmentation as a way of disproving the neoclassical theory. However, it is also possible for the two ideas to complement each other. Specifically, it might be the case that what prevents poor people from getting good jobs are indeed conventional productivity variables, such as education and experience, as neoclassical theory suggests. But even so, the labor market might still be usefully characterized as segmented if the wage functions of those in the primary labor market and those in the secondary labor market are different enough. We use segmentation theory here to characterize inequality in the Israeli labor market, rather than to attempt to prove the applicability of one theory over the other to Israel.

Knowing whether the labor market is segmented is important for labor market policy, because a given set of policy instruments might not affect all workers equally if the labor market is segmented, or might not even affect certain workers at all. For example, the focus on schooling and training as tools to increase wages of the less-skilled may not be adequate if the returns to human capital are low in the less-skilled segment of the labor force. Hence, it is not only important to know whether the labor market is segmented; it is even more important to characterize the different segments. This is the purpose of this paper.

In the next section we review the literature on labor market segmentation. After that we present the methodology to be used in the empirical analysis. The data are described next, and then we present the empirical results. The final section summarizes the findings and discusses their implications.

Literature on labor market segmentation

Dickens and Lang (1985) lay out the two hypotheses that, if satisfied, would demonstrate the existence of a segmented labor market and cast doubt on the neoclassical model. The first is that the labor market can be described well by the dual market typology; that is, that most jobs resemble either "good" (primary sector) jobs or "bad" (secondary) jobs, rather than having a continuum of jobs and workers in the economy. The second is that rationing of primary sector jobs takes place. They develop a switching regressions technique, which makes separate inferences for different labor market segments using a separate sorting equation to predict each observation’s propensity to be in one segment or another. With this method, they ask whether two wage equations fit the data significantly better than one and test for the existence of barriers to enter the primary sector. Using data from the thirteenth wave of the American Panel Study of Income Dynamics, they find empirical support for both hypotheses underlying labor market segmentation. They find that workers in standard Metropolitan Statistical Areas (that is, workers who live near cities), married workers, more educated workers, and whites are less likely to be in the secondary sector.

Roig (1999) applies the method suggested by Dickens and Lang (1985) to Spanish data and also finds evidence of a dual structure of the Spanish labor market, along with evidence that some workers find it much easier to find a job in the primary sector than others do. Paihle (2003) applies this method to data from the formerly planned Central European countries of the former Czechoslovakia, Poland, and Hungary and also finds significantly different coefficients for two different sectors of the economy, with some workers having more access to the primary sector – that is, the sector with higher returns to human capital – than others.

Sousa-Poza (2004) uses three different methods to test for segmentation in the Swiss labor market. First, he performs cluster analysis on Swiss Labor Force Survey (SLFS) data from 2000 and does not find evidence of a secondary labor market in Switzerland. He then uses the Dickens and Lang (1985) switching model on the same SLFS data and does find evidence of segmentation. Finally, he applies a discrete choice model to the first three waves of the Swiss Household Panel (SHP) to measure low-wage mobility, estimating the probability of those in low-wage jobs in 1999 moving to high-wage jobs by 2001. He finds that a higher percentage of workers remain in low-wage jobs than would be expected if there were not barriers to the primary sector. People with low education were more likely to remain in low-wage jobs, and people in large firms were more likely to move to the high-wage market, but otherwise, reasons for wage mobility were not explained by observed characteristics of workers or industries.

Pittau and Zelli (2006) use a mixture regression model, which allows the distribution of a random variable to be expressed as a mixture of several different distributions (see also Laird, 1978; Heckman and Singer, 1984). As opposed to the switching regression of Dickens and Lang (1985), it does not require each observation to strictly belong to one segment, but allows each observation to belong to each segment with some probability. It also has the advantage that the number of segments can be changed easily, and inference can be made on the appropriate number of segments. They find that a mixture of three or four normal segments fits the data well for Italian per-capita income distributions. Battisti (2013) uses Italian data to find evidence of multiple wage equations and further evidence that movement into the primary sector is more difficult than general movement among job positions. Pittau et al. (2010) used mixture models to study income distribution in a panel of countries. Mixture regression has been used to find segmentation in heterogeneous populations in other fields, including marketing (Wedel et al., 1993), health economics (Deb and Trivedi, 1997; Conway and Deb, 2005; Deb et al., 2011), and transportation (Park and Lord, 2009).

Methodology

In this section we describe the mixture regression procedure, in which a wage regression is used to characterize the data, and observations are allowed to have a probability distribution over a number of different wage regressions.

Suppose that the labor market in composed of S different segments with wages in each segment having a unique distribution, and assume that log-wages are normally distributed with segment-specific means and variances. Therefore, the density function of a worker's log-wage conditional on belonging to segment s is:

$$(1) \qquad f(w_i \mid s; \beta_s, \sigma_s) = (2\pi\sigma_s^2)^{-\frac{1}{2}} \exp\left(-\frac{[w_i - \sum_j \beta_{j,s} x_{i,j}]^2}{2\sigma_s^2} \right)$$

where w_i is the i^{th} individual's log-wage, $x_{i,j}$ is the j^{th} (j=1,…,J) explanatory variable of individual i, $\beta_{j,s}$ is the corresponding regression coefficient in the s^{th} segment, and σ^2_s is the residual variance of the s^{th} segment. While each individual belongs to one and only one segment, the assignment of individuals to segments is unknown. Let μ_s be the weight attached to segment s. These weights can be interpreted as the unconditional probabilities that an individual belongs to each segment. Then we get the density function for a "finite mixture" with S segments (Pittau and Zelli, 2006):

$$(2) \qquad f(w_i \mid \boldsymbol{\beta}\ ,\boldsymbol{\sigma}\ ) = \sum_S \mu_s (2\pi\sigma_s^2)^{-\frac{1}{2}} \exp\left(-\frac{[w_i - \sum_j \beta_{j,s} x_{i,j}]^2}{2\sigma_s^2} \right)$$

where $\boldsymbol{\beta}$ is a matrix consisting of all regression coefficients of all segments and $\boldsymbol{\sigma}$ is a vector consisting of all standard deviations. Flachaire and Nuñez (2007) note that under regularity conditions, any probability density can be consistently estimated by a mixture of normal densities, hence the normality assumption is not as restrictive as one may intuitively think.

Given n independent observations that are distributed as (2), the likelihood function of the mixture model is written as:

$$(3) \qquad L = \prod_{i=1}^{n} \left[\sum_S \mu_s (2\pi\sigma_s^2)^{-\frac{1}{2}} \exp\left(-\frac{[w_i - \sum_j \beta_{j,s} x_{i,j}]^2}{2\sigma_s^2} \right) \right]$$

This likelihood function can be maximized using the EM algorithm (Wedel et al., 1993; Pittau et al., 2010; Battisti, 2013). Given the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ (the vector of segment weights), it is possible to compute the posterior probability that observation *i* belongs to segment *s*, with Bayes' theorem:

$$(4) \qquad p_{i,s} = \frac{\hat{\mu}_s f_{i,s}(w_i \mid x_{i,j=1...J}, \hat{\sigma}_s^2, \hat{\beta}_{j=1...J,s})}{\sum_{s=1}^{S} \hat{\mu}_s f_{i,s}(w_i \mid x_{i,j=1...J}, \hat{\sigma}_s^2, \hat{\beta}_{j=1...J,s})}$$

The interpretation of (4) is the following: The probability that observation *i* belongs to segment *s* equals the probability that any observation is in segment *s* times the probability of obtaining the parameter estimates we obtained, were this observation truly in segment *s*, divided by the probability of obtaining the parameter estimates we obtained, for any segment.[1]

These posterior probabilities could then used to assign each observation to the segment for which the posterior probability of membership is highest, in order to compare the segments in terms of their observed characteristics.

Data

Data for this analysis are taken from the Combined 2010 Income Survey of Israel's Central Bureau of Statistics (CBS). The file combines data from two surveys so that the combined sample is representative of the Israeli population. [2] In both surveys, every adult living in a sampled household is questioned. In 2010, the two surveys combined to provide data on 36,331 individuals in 15,171 households. Of these, 17,019 were wage-earning

---

[1] Bayes' Theorem is most simply expressed in the equation:

$$\Pr(A \mid X) = \frac{\Pr(X \mid A)\Pr(A)}{\Pr(X)},$$

where Pr (A|X) is the probability of the event (A) given a positive test (X), Pr(X|A) is the probability of a positive test, given the event, Pr(A) is the unconditional probability of the event, and Pr(X) is the unconditional probability of a positive test. In our case, the "event" is observation *i* belonging to segment *s* and the "positive test" is the set of parameter estimates. The following table explains the corresponding notation of equation (4) and Bayes' Theorem:

General notation	Notation in equation (4)	Explanation
$\Pr(A \mid X)$	$p_{i,s}$	Probability that observation *i* is in segment *s* (implied: given the parameter estimates obtained). This is the posterior probability.
$\Pr(X \mid A)$	$f_{i,s}(w_i \mid x_{i,j=1...J}, \hat{\sigma}_s^2, \hat{\beta}_{j=1...J,s})$	Probability that the parameter estimates obtained would be obtained, given that *i* belongs to *s*
$\Pr(A)$	$\hat{\mu}_s$	Unconditional (prior) probability that any observation belongs to *s*
$\Pr(X)$	$\sum_{s=1}^{k} \hat{\mu}_s f_{i,s}(w_i \mid x_{i,j=1...J}, \hat{\sigma}_s^2, \hat{\beta}_{j=1...J,s})$	Probability of obtaining the parameter estimates obtained, over all segments.

[2] The first is an income survey that is conducted along with the annual Labor Force Survey, which samples households throughout the country. The second is the annual Household Expenditure Survey, which also samples households throughout the country and asks about income (State of Israel, 2012).

employees, and complete information was available for 16,897.[3] The average hourly wage in this sample was 46 NIS (Table 1).[4]

In addition to the wage, the following variables are used in the empirical analysis. Their descriptive statistics are shown in Table 1.

- Gender: We would expect female workers to be overrepresented in the secondary labor market, since it is known that women, on average, earn lower wages than men do, for a variety of reasons, including choices to work in less remunerative occupations and industries, fewer years of work experience, and discrimination.[5]
- Age: Since we do not have direct data on workplace experience, age will act a proxy, with older workers likely to have more experience and more access to the primary sector.
- Years of schooling: The secondary labor market is usually characterized by low levels of education (Sousa-Poza 2004).
- Yeshiva status and yeshiva household status: We use indicators for whether the individual studied in a yeshiva (rabbinic seminary for adults), and whether the individual lives in the same household as someone who studied in a yeshiva, as indicators of Haredi (ultra-Orthodox) status.[6] We expect former yeshiva students to have less access to the primary labor market, given the low overall levels of employment among Haredi Jews (Kimhi 2012).
- Nationality: In Israel, we expect Arabs to be underrepresented in the primary sector, as they might face discrimination, lower quality of education, and other barriers to the primary job market (Haberfield and Cohen 2007). The nationality and yeshiva household status variables thus delineate three social categories, or sectors, of the Israeli labor market: Arabs, Haredi Jews, and non-Haredi Jews.
- Ethnicity: In the Israeli context, this variable serves as a finer version of the "nationality" variable, with past research showing labor market disparities between the different ethnic groups (Haberfield and Cohen 2007). Hence, instead of breaking up the sample into Jews and Arabs, we break it up into five ethnic groups:
    - Arabs;
    - Sefaradi Jews (Sefaradim): Jews who themselves immigrated to Israel, or whose fathers immigrated to Israel, from Asia or Africa;
    - Ashkenazi Jews (Ashkenazim): Jews who themselves immigrated to Israel, or whose fathers immigrated to Israel, from Europe (excluding the Former Soviet Union), the Americas, or Oceania;
    - Israeli Jews: Jews who themselves were born in Israel or British mandatory Palestine, <u>and</u> whose fathers were also born in Israel or British mandatory Palestine; and
    - Immigrants from the Former Soviet Union, who are not counted as Sefaradi or Ashkenazi Jews. Most of these are fairly recent immigrants that, although showing remarkable labor market assimilation, are more likely to be in the secondary labor market.
- Periphery status / Geographic location: The CBS assigned each household a number, on a scale of 1 to 5, indicating how far the family lived from the center of the country. A number 1 indicates a household considered "very peripheral," and a 5 indicates "very

---

[3] Three observations were dropped because they lacked wage data, 13 were dropped because they lacked data on educational background, and 106 were dropped because they lacked data on ethnic background. 607 workers did not report industry of occupation, and those were excluded when necessary.

[4] Hourly wage was calculated as ([monthly wage]*12) / ([weekly hours worked]*52).

[5] The literature on the determinants of the wage differences between men and women is immense. See, for example, Blau and Kahn (1997).

[6] This method of identifying Haredi Jews aligns with Approach B in Fridman et al. (2011).

central." Following Krueger and Summers (1988), we hypothesize that those in the periphery will have less access to primary sector jobs.

- Industry and occupation: Krueger and Summers (1988) show that the industry a person works in and their occupation help determine their wage, independent of the person's human capital. Past studies of segmentation in other labor markets, such as Sousa-Poza (2004), show industry playing a central role in segmentation. We would expect professional and managerial occupations to garner higher wages than other occupations. However, it is difficult to hypothesize which industries would provide more or fewer jobs in the primary sector. The CBS data include a categorical variable indicating the worker's industry that takes one of fifteen values, including "industry unknown." For regressions that included dummy variables for occupation and industry, observations that had "industry unknown" or "occupation unknown" were dropped, leaving a total of 16,290 observations.

Empirical Results

Cluster analysis

Before going to the regression framework, we perform a somewhat descriptive diagnosis of potential labor market segmentation. For this purpose we conduct cluster analysis of the data. The *k*-means cluster procedure is an iterative procedure for splitting a sample into *k* groups that are as homogenous as possible, so that the ratio of the distance between observations of the same cluster and the distance between clusters is minimized, where the multi-dimensional distance can be computed using any number of variables. This procedure does not distinguish between dependent variables and explanatory variables. The algorithm is composed of the following steps:

1. Choose initial cluster centers (could be arbitrary but should be far apart);
2. Assign each subject to the 'nearest' cluster, defined in terms of the distance to the centroid;
3. Find the centroids of the clusters that have been formed;
4. Repeat steps (2) and (3) until the centroids remain relatively stable.

It is recommended to convert continuous variables into their z-scores so that no single variable dominates the distance measure (Norusis, 2011).

The cluster analysis was executed using the Quick Cluster procedure in SPSS. The results are in Table 2.[7] Cluster analysis with k=2 yielded a high-wage group, encompassing about a third of the workers, with almost two and a half times the average wage of the low-wage group (76 shekels per hour compared to 30 shekels per hour). Age is highly correlated with wage, with the high-wage cluster being 11 years older, on average, than the low-wage cluster. The same is true for schooling, with the high-wage cluster having about 4 more years of schooling than the low-wage cluster. Contrary to what one might expect, female workers and immigrants from the Former Soviet Union were about equally represented in each group. Native Israelis (those whose fathers were born in Israel) and Sefaradim were slightly overrepresented in the low-wage cluster, with native Israelis making up 25% of the low-wage cluster but just 19% of the high-wage cluster, and Sefaradim making up 31% of the low-wage cluster and 22% of the high-wage cluster. The clustering of Ashkenazim, and especially of Arabs, is sharper. Ashkenazim made up 31% of the high-wage cluster and 9% of the low-

[7] Because k-means cluster analysis is sensitive to outliers (Norusis 2011), the analysis was then repeated after excluding 225 observations that had either implausibly low reported wages of less than five shekels per hour or extremely high wages of over 300 shekels per hour. The distribution of variables by assigned cluster did not look very different with and without these outliers, suggesting that their presence does not affect the results considerably.

wage cluster, while Arabs made up just 4% of the high-wage cluster and 14% of the low-wage cluster.[8] Yeshiva graduates were 2% of the high-wage cluster and 1% of the low-wage cluster. This is a surprising result that might be due to multicollinearity.

Workers in the high-wage cluster were slightly more likely to be in the very central region of the country, while more peripheral and intermediate regions were over-represented in the low-wage cluster. Industries with a higher representation in the high-wage cluster included utilities, finance, real estate, public administration, education and health. Industries with a higher representation in the low-wage cluster included agriculture, manufacturing, construction, commerce, hospitality, transportation, and domestic work. Occupations with a higher concentration in the high-wage cluster included academic, technical and managerial occupations (by a large margin). Occupations with a higher concentration in the low-wage cluster included Clerical, sales, agricultural, professional and unskilled occupations.

Cluster analysis with k=3 produced a similar pattern of results but with some sharper distinctions. Clustering into three groups produced high-wage, medium-wage, and low-wage groups, but while the high-wage group was far apart from the medium-wage group, averaging 123 shekels per hour, the medium-wage and low-wage groups were not that much different, averaging 44 and 30 shekels per hour, respectively. Age differentials between clusters mimic the wage differentials. Schooling is also found to be correlated with wage, but in this case the schooling differential between the medium-wage and low-wage clusters is much larger than the schooling differential between the high-wage cluster and the medium-wage cluster. As opposed to the 2-cluster case, gender seems to be a factor here, with females concentrated mostly in the medium-wage cluster (64%), and more in the low-wage cluster (46%) than in the high-wage cluster (36%). Arabs, on the other hand, are over-represented in the low-wage cluster (16%, compared to 6% and 3% in the medium-wage and high-wage clusters, respectively). Ashkenazim represent a larger share of the high-wage cluster (39% compared to 21% and 8% in the medium-wage and low-wage clusters, respectively), and immigrants from the Former Soviet Union represent a larger share of the medium-wage cluster (29%, compared with 19% and 14% in the low-wage and high-wage clusters, respectively). The distribution of the periphery status indicators did not change qualitatively from the two-cluster results, but there were a number of changes related to industry and occupation. In particular, the education and health industries are over-represented in the medium-wage cluster, and the same is true for technical occupations.

The cluster analysis results show that the Israeli labor market is indeed potentially segmented. We identified a number of worker attributes that are correlated with segments characterized by different levels of hourly wage, including human capital, demographics, location, industry and occupation. We now want to establish the dependence of wages on these attributes, allowing for segmentation that is manifested as different returns to these attributes in different segments.

Wage regressions

Before moving to the results of the mixture regression model, we present OLS results of the wage regression (Table 3). The dependent variable in all regressions is the log of hourly wage. Column 1 is the basic specification, and we can see positive and diminishing returns to age, positive returns to schooling, and wage disadvantages for females, Arabs and Haredim. After controlling for location, industry and occupation effects (column 2), the returns to education are halved (from 7.7% to 3.5%), while the wage disadvantages of

[8] The cluster analysis was run both with and without the ethnic background variables, and using both the individual's own place of birth and the father's place of birth as the determining factor in determining an individual's ethnicity. An observation's assigned cluster ended up being the same for more than 90% of cases, regardless of whether and how the ethnic background variable was used.

females, Arabs and Haredim also decline. In column 3 we interact gender and sector. This shows that the largest wage disadvantage is among Haredi males, while no statistically significant disadvantage was found among Haredi females. Arab males also suffer a much larger wage disadvantage than Arab females. Only among the non-Haredi Jewish majority, females earn less than males, other things equal. In column 4 we interact schooling with gender, and find slightly larger returns on education for females. At the same time, the average wage penalty of females becomes larger, indicating that the less-educated females are particularly disadvantaged. In column 5 we interact schooling with sector. We find practically zero returns to schooling among Arabs, while Haredim have negative returns to schooling.[9] The regression in column 6 shows full interactions of gender, sector and schooling. We now see quite similar returns to schooling for males and females among the non-Haredi Jewish majority, positive but much lower returns to schooling among Arab females, negative returns to schooling for Arab and Haredi males, and zero returns to schooling for Haredi females. The wage penalty for females is much larger for Arab females than for Jewish females, while there is no wage penalty for Haredi females. Arab males do not earn significantly less than non-Haredi Jewish males, while Haredi males still earn more than other males, other things equal. Column 7 shows that the qualitative results are unchanged after controlling for location, industry and occupation. However, the returns to education are still halved compared to column 6, regardless of being positive or negative, and the negative returns to schooling among Arab males becomes statistically insignificant. This implies that much of the educational wage differentials are between rather within groups of workers defined by location, industry and occupation.

The estimated wage equations gave reasonable results concerning wage impacts of human capital and demographics. As the existing literature (e.g., Kimhi 2012) and a cursory look at the economy show, those with more education and experience, those in high-skilled occupations and high-value-added industries, those living in the center of the country, non-Haredi Jews, and men, have higher wages than others in the Israeli labor market. The question is whether there are systematic latent differences in wages that cut across population groups, and are therefore not accounted for in the regression analysis. If these unobserved differences are correlated with the observed explanatory variables, the estimated wage impacts could be biased. To account for potential wage differences as well as potential differences in returns to observable attributes across latent worker classes, we use the mixture regression analysis, which allows for different wage equations in different latent segments of the work force.

Mixture models require a priori choice of the number of segments, each having a specific wage equation. Thus, each specification was estimated with two and three segments and the explanatory power of the two-segment and three-segment mixtures was compared, both in terms of formal statistical criteria – the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) – and in terms of which mixtures had more intuitively explainable results (Battisti 2013). We used the command *fmm*, for finite mixture modeling, in Stata 11 (Deb 2007) to run the models, and the related command *fmmlc* for post-estimation (Luedicke 2011).

Mixture regression results

Table 4 includes the mixture regression results for the two-segment and three-segment specifications. In the two-segment specification, less than 18% of the sample observations are assigned to segment 1 and the remaining 82% to segment 2. At the bottom of the table we

[9] Kimhi and Sandel (2011) found that per-capita household income is negatively related to years of schooling of Haredi heads of household.

report predicted values of log-wages, computed at the sample means. It can be seen that segment 2 is the high-wage segment. The two wage equations differ substantially from each other. For example, the female wage penalty is statistically significant only in segment 2. The same is true for the returns to schooling among Arabs and for the wage penalty in the periphery. On the other hand, the occupational wage penalty of unskilled workers (the excluded group) is larger in segment 1, and the same is true for the industrial wage penalty in agriculture, hospitality, education and health, compared to manufacturing. The standard deviation of wages in segment 1 is more than twice as large as the standard deviation of wages in segment 2.

The segment 1 wage equation in the three-segment specification is quite similar to the segment 1 wage equation in the two-segment specification, both in terms of its size (17% of the sample) and in terms of the coefficients. It still emerges as the low-wage segment, and its wage variability is large. Of the remaining two segments, segment 2, with 34% of the sample observations, has a lower predicted wage than segment 3, with 49% of the sample observations. The coefficients of the wage equations of segments 2 and 3 are quite different. A female wage penalty for Jewish females exists only in segment 3, while for Arab females it is statistically significant only in segment 2. The male Arab wage penalty is statistically significant only in segment 3. The returns to schooling for Jewish males and females are much larger in segment 3, while the returns to schooling for Arab females are significant only in segment 2. Wages in sales and agricultural occupations are higher than wages in unskilled occupations only in segment 3, while wages in professional occupations are higher than wages in unskilled occupations only in segment 2. The wage penalty in agriculture, construction, commerce, hospitality, education, health, community and domestic services compared to manufacturing is significant only in segment 3 (wages in education and in domestic services in segment 2 are actually higher than in manufacturing), while the wage penalty in real estate is significant only in segment 2.

Judging by the values of AIC, BIC and likelihood, the three-segment specification is preferred over the two-segment specification. The fact that the majority segment in the two-segment specification is now split into two different segments adds flexibility to the model and allows different effects of explanatory variables in segments 2 and 3. For example, the male Arab wage penalty is not significant in segment 2 of the two-segment specification, while it is significant in segment 3. On the other hand, the schooling wage premium of Arab males is significant in segment 2 of the two-segment specification, while it is insignificant in segments 2 and 3 of the three-segment specification and even significantly negative in segment 1.

Simulated segments

The posterior probability of each observation to belong to each segment is computed using (4), and each observation is then assigned to the segment for which the posterior probability of belonging is highest. Thus, the sample is divided into "simulated segments". Table 5 shows the variable means in each simulated segment. Beginning with the two-segment specification, it can be seen that average wage is higher in segment 1, as opposed to the prediction using the overall sample means. This implies that the population composition of the segment leads to a higher predicted wage. In fact, table 5 shows that segment 1 includes considerably fewer Arabs and slightly more educated workers. The fact that Arabs have lower schooling, on average (Shavit and Bronstein, 2011), and that the return to schooling in segment 1 is statistically significant only for Jewish workers, explains the higher predicted wage in segment 1. Note that the fraction of observations assigned to segment 1, 8%, is much lower than the estimated relative size of the segment, 18% (table 4), and this casts some doubt about the adequacy of this method of assignment of observations to segments. No other considerable differences are observed between the simulated-segment means of segments 1 and 2.

In the three-segment specification, segment 1, now including 9% of the sample, is no longer the high-wage segment. Its average wage is much lower than the average wage in segment 3 and not so much higher than the average wage in segment 2. Segment 1 has fewer Arab workers, but other than that, the variable means in all three segments are not very different from each other. This may lead to the conclusion that the Israeli labor market is not so much segmented by observable worker attributes but rather by unobservable factors.

However, there is still concern regarding the deviation of the population fractions of the assigned segments from their estimated values. In fact, the way the assigned segments are constructed may introduce assignment errors in the sense that, in the case of the two-segment specification, observations with posterior probabilities of 0.99 and 0.51 are grouped together in the same assigned segment, despite the fact that for the second observation there is an almost equal probability that it actually belongs to the other segment. Consequently, we propose an alternative method of generating assigned segments, based on weighting the observations by posterior probabilities rather than assigning by the highest probability. Specifically, we include all observations in each of the simulated segments, where each observation is weighted by the posterior probability of the relevant segment, so that the sum of weights within each simulated segment is equal to the estimated unconditional probabilities μ_s.

The variable means of the "weighted simulated segments" are reported in table 6. Not surprisingly, the differences in the variable means across the segments are in general smaller than those of the assigned segments in table 5. In addition, in the three-segment specification, the average wage in the intermediate wage segment (segment 1) is now closer to the average wage of the high-wage segment (segment 3) than to the average wage of the low-wage segment (segment 2). Recall that segment 1 is not only the smallest among the three. It also has a considerably larger wage variability than the other two segments. This leads to the conclusion that segment 1 of the hired labor force includes both high and low wages that do not fit well the wage equations of either the low-wage segment or the high-wage segment.

The 3-segment specification of the mixture regression model helps, therefore, to "purge" the observations with wages that are difficult to assign based on observable characteristics. Segments 2 and 3 represent, in this case, low-wage and high-wage employees, respectively. The segment-specific means in table 6 show the contribution of observable characteristics to the wage disparity. Although they are quantitatively small, the differences between the means indicate that age and schooling are associated with a higher wage. Females are equally represented in the two segments, while Arabs are over-represented in the low-wage segment. Living in the periphery is associated with lower wages, and the same is true for working in construction, commerce, hospitality, health, and domestic services sectors. On the other hand, working in the finance, real estate, public services, and education sectors is associated with higher wages. Academic, technical, managerial and clerical occupations imply higher wages, while sales, professional and unskilled occupations imply lower wages.

Summary and discussion

Our empirical analysis has shown that the mixture regression model can identify different segments in the Israeli labor market. The three-segment model differentiated between a low-wage segment, a high-wage segment, and a third segment with both high and low wages that cannot be assigned to either of the other two segments. One interpretation of this result is that the model allows "purging" of "noisy" observations, those with unexpectedly low or high wages, conditional on their observed characteristics. Given this interpretation, the three-segment model allows a clearer distinction between low-wage and high-wage jobs.

Applying the common method of assigning observations to simulated segments using the maximal estimated posterior probability, we found that the assigned sample shares of the simulated segments are considerably different than their estimated population shares. This result leads to the conclusion that this method might lead to systematic classification errors. Instead, we propose weighting all observations by their estimated posterior probabilities and including all weighted observations in each of the simulated segments. This leads, by definition, to simulated segments with sample shares that are identical to the estimated sample shares. Application of this classification method yielded simulated segments that are not dramatically different than those obtained using the conventional method, but still lead to some different insights. For example, Arabs were equally likely to be assigned to the high-wage and low-wage assigned segments, while they were more likely to be in the low-wage weighted segment than in the high-wage one. The same is true for residents of the periphery.

Altogether, the weighted simulation method yielded quantitatively small but qualitatively reasonable differences between the attributes of workers in the low-wage segment and the high-wage segment. The differences between the two segments in terms of their wage equation coefficients, or the returns to those attributes, were much larger. For example, Jewish females and Arab males suffer considerable wage penalties in the high-wage segment, while Arab females suffer wage penalties in the low-wage segment. Returns to schooling are considerably higher for Jewish workers in the high-wage segment, while the contrary is true for Arab females. Altogether, the results indicate that much of the wage disparities in Israel are due to unobserved factors rather than to observable characteristics.

The results lead to some interesting policy-relevant findings on the links among schooling, ethnic minorities, and wages. Male Arabs, for example, do not enjoy returns to schooling, regardless of their wage segment. Hence, they can earn at most as much as uneducated male Jewish workers in the low-wage segment, but much lower than any male Jewish worker in the high-wage segment. Female Jewish workers can earn as much as male Jewish workers in the low-wage segment, but much lower than an equally-educated and otherwise identical male Jewish worker in the high-wage segment. Female Arab workers earn less than any other worker in the low-wage segment if they have no schooling, but can compensate for that disadvantage if the acquire sufficient schooling. If they happen to be in the high-wage segment, on the other hand, additional schooling seems to be worthless for them.

This last result is particularly interesting and relevant for public policy, because it implies that while schooling is advantageous for female Arab workers, it is not sufficient to lift them out of the low-wage trap. The challenge of policy, in this case, is to find ways to integrate female Arabs into the high-wage segment, and at the same time find ways to make their schooling valuable in that market segment. This is an example of the policy relevance of the mixture regression model of the labor market.

One remaining challenge that deserves further research is to understand how observationally-equivalent workers are assigned to the different segments. In the Israeli case, it is common to talk about “connected” and “unconnected” workers. The former include public employees, who perhaps do not enjoy high wages but are compensated by extreme job security arrangements and fringe benefits, workers in public utilities and banks that have extremely powerful unions, as well as others whose parents or friends hold influential positions. It is an even more important challenge of policy to make sure that the ability to find a job that pays well will be based principally if not solely on merit.

References

Atkinson, Anthony B., and Andrea Brandolini (2011). *On the Identification of the "Middle Class"*. Society for the Study of Economic Inequality Working Paper No. 2011-217. Available at http://www.ecineq.org/milano/wp/ecineq2011-217.pdf (retrieved 10 July 2013).

Battisti, Michele (2013). "Reassessing Segmentation in the Labour Market: An Application for Italy 1995-2004," *Bulletin of Economic Research* 65(s1), pp. s38-s65. Available at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8586.2011.00415.x/pdf (retrieved 10 July 2013).

Birdsall Nancy, Carol Graham, and Stefano Pettinato (2000). *Stuck in The Tunnel: Is Globalization Muddling The Middle Class*? Brookings Institution, Center on Social and Economic Dynamics Working Paper No. 14. Available at http://www.brookings.edu/es/dynamics/papers/middleclass/middleclass.pdf (retrieved10 July 2013).

Blau, Francine D., and Lawrence M. Kahn (1997). "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics* 15(1) Part 1, pp. 1-42.

Burkhauser, Richard V., Amy Crews Cutts, Mary C. Daly, and Stephen P. Jenkins (1999). "Testing the Significance of Income Distribution Changes over the 1980s Business Cycle: A Cross-National Comparison." *Journal of Applied Econometrics* 14(3), pp. 253-272.

State of Israel, Central Bureau of Statistics (2012). *Income Survey 2010*. Publication No. 1479. Available at http://www1.cbs.gov.il/publications12/1479/pdf/e_print.pdf (accessed July 15th, 2012).

Cohen, Yinon, Yitzchak Haberfeld, Tali Kristal, and Guy Mundlak (2007). "The State of Organized Labor in Israel," *Journal of Labor Research* 28(2), pp. 255-273.

Conway, Karen Smith, and Partha Deb (2005). "Is Prenatal Care Really Ineffective? Or, Is the 'Devil' in the Distribution?" *Journal of Health Economics* 24(3), pp. 489-513.

De Sarbo, Wayne S., and William L. Cron (1988). "A Maximum Likelihood Methodology for Clusterwise Linear Regression," *Journal of Classification* 5, pp. 249-282.

Deb, Partha (2007). "FMM: Stata module to estimate finite mixture models," Statistical Software Components, S456895, Boston College Department of Economics, revised 12 Feb 2012. Available at http://ideas.repec.org/c/boc/bocode/s456895.html (retrieved 29 March 2012).

Deb, Partha, William T. Gallo, Padmaja Ayyagari, Jason M. Fletcher, and Jody L. Sindelar (2011). "The Effect of Job Loss on Overweight and Drinking." *Journal of Health Economics* 30(2), pp. 317-327.

Deb, Partha, and Pravin K. Trivedi (1997). "Demand for Medical Care by the Elderly: A Finite Mixture Approach." *Journal of Applied Econometrics* 12, pp. 313-336.

Dickens, William T., and Kevin Lang (1985). "A Test of Dual Labor Market Theory," *The American Economic Review* 75(4), pp. 792-805.

Esteban, Joan-Maria, and Debraj Ray (1994). “On the Measurement of Polarization.” *Econometrica* 62(4), pp. 819-851.

Flachaire, Emmanuel, and Olivier Nuñez (2007). “Estimation of the Income Distribution and Detection of Subpopulations: An Explanatory Model.” *Computational Statistics & Data Analysis* 51(1), pp. 3368-3380.

Fridman, Israela, Nava Shau-Mena, Nir Fogel, Dmitri Romanov, Dan Amedey, Mark Feldman, Ruth Sehayek, Gustavo Schifris, and Haim Portnoy (2011). "Measurement and Estimates of the Population of Ultra-orthodox Jews" (Hebrew), Israel Central Bureau of Statistics. Available at http://www1.cbs.gov.il/www/publications/tec25.pdf (retrieved 31 May 2012).

Haberfeld, Yitzchak, and Yinon Cohen (2007). "Gender, Ethnic, and National Earnings Gaps in Israel: The Role of Rising Inequality," *Social Science Research* 36(2), pp. 654-672.

Heckman, James J., and Burton Singer (1984). “A Method of Minimizing the Distributional Impact in Econometric Models for Duration Data.” *Econometrica* 52, pp. 271-320.

Israel Central Bureau of Statistics (CBS) website and publications on Income Survey. http://cbs.gov.il/.

Kimhi, Ayal (2012). “Labor Market Trends: Employment Rate and Wage Disparities.” In Ben-David, Dan (Ed.). *State of the Nation Report: Society, Economy and Policy in Israel 2011-2012*. Jerusalem: Taub Center for Social Policy Studies in Israel.

Kimhi, Ayal, and Moran Sandel (2011). *Religious Schooling, Secular Schooling, and Household Income Inequality in Israel*. AlmaLaurea Working Papers No. 29.

Krueger, Alan B, and Lawrence H. Summers (1988). "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica* 56(2), pp. 259-293.

Laird, N. (1978). “Nonparametric Maximum Likelihood Estimation of a Mixing Distribution.” *Journal of the American Statistical Association* 73, pp. 805-811.

Lubell, Maayan (2011). "Ultra-religious schools test Israel's high-tech future," 22 December. Available at http://www.reuters.com/article/2011/12/22/us-israel-ultraorthodox-economy-idUSTRE7BL0J720111222 (retrieved 31 May 2012).

Luedicke, Joerg (2011). "FMMLC: Stata module for postestimation with finite mixture models," Statistical Software Components, S457291, Boston College Department of Economics. Available at http://ideas.repec.org/c/boc/bocode/s457291.html (retrieved 29 March 2012).

Norusis, Marija (2011). "Cluster Analysis," in *IBM SPSS Statistics – Statistical Procedures Companion*, version 19. Available at http://norusis.com/pdf/SPC_v19.pdf (retrieved 07 March 2012).

Pailhe, Ariane (2003). "Labour Market Segmentation in Central Europe during the First Years of Transition," *Labour* 17(1), pp. 127-152.

Park, Byung-Jung, and Dominique Lord (2009). "Application of Finite Mixture Models for Vehicle Crash Data Analysis." *Accident Analysis and Prevention* 41(4), pp. 683-691.

Pittau, M. Grazia, and Roberto Zelli (2006). "Trends in Income Distribution in Italy: A Non-Parametric and a Semi-Parametric Analysis." *Journal of Income Distribution* 15, pp. 90-118.

Pittau, M. Grazia, Roberto Zelli, and Paul A. Johnson (2010). "Mixture Models, Convergence Clubs, and Polarization." *Review of Income and Wealth* 56(1), pp. 102-122.

Roig, Ana Huguet (1999). "Testing Spanish Labour Market Segmentation: An Unknown-Regime Approach," *Applied Economics* 31, pp. 293-305.

Shavit, Yossi, and Vicki Bronstein (2011). "Education Reform and Narrowing Educational Gaps in Israel." In Ben-David, Dan (Ed.). *State of the Nation Report: Society, Economy and Policy in Israel 2010*. Jerusalem: Taub Center for Social Policy Studies in Israel.

Sousa-Poza, Alfonso (2004). "Is the Swiss Labor Market Segmented? An Analysis Using Alternative Approaches", *Labour* 18(1), pp. 131-161.

Taub Center for Social Policy Studies in Israel (2011). "Non-employment amongst ultra-Orthodox men has risen by 200% during the last 30 years," press release. Available at http://taubcenter.org.il/tauborgilwp/wp-content/uploads/PR_E2010_State_of_Nation_Report.pdf (retrieved 08 March 2012).

Ravallion, Martin (2010). "The Developing World's Bulging (but Vulnerable) Middle Class." *World Development* 38(4), pp. 445-454.

Wedel, M., Desarbo, W.S., Bult, J.R., Ramaswamy, V. (1993). "A Latent Class Poisson Regression Model for Heterogeneous Count Data." *Journal of Applied Econometrics* 8, 397-411.

Table 1: Descriptive statistics of the sample

Category	Variable	Label	Mean
	Hourly wage	Hourly Wage	46.0
	Ln(hourly wage)	Log of hourly wage	3.6
	Schooling	Years of schooling	13.9
	Age	Age	40.3
	Peripherality	Peripherality index	3.9
Sector	Yeshiva	Studied in yeshiva (1=yes)	1.1%
Sector	Haredi	In household of yeshiva student	1.7%
Sector	Arab	Arab	10.7%
Gender	Female	Female	50.3%
Sector/Gender	Female*Majority	Female non-Haredi Jew	46.6%
Sector/Gender	Female*Haredi	Female Yeshiva household member	0.7%
Sector/Gender	Female*Arab	Female Arab	3.0%
Ethnicity	Ashkenazi	Individual or father Ashkenazi	16.6%
Ethnicity	Sefaradi	Individual or father Sefaradi	27.7%
Ethnicity	Native Israeli	Individual and father Israeli-born	23.4%
Ethnicity	USSR	Immigrant from former USSR	21.6%
Location	Very peripheral	Lives in distant periphery (1=yes)	3.7%
Location	Peripheral	Lives in periphery	8.2%
Location	Intermediate	Lives between center and periphery	29.5%
Location	Central	Lives in center	14.3%
Location	Very central	Lives in a close center	44.3%
Industry	Agriculture	Agriculture	1.1%
Industry	Manufacturing	Manufacturing	15.1%
Industry	Utilities	Utilities	0.7%
Industry	Construction	Construction	4.7%
Industry	Commerce	Motors, retail, wholesale	12.0%
Industry	Hospitality	Hospitality	4.6%
Industry	Transportation	Transportation	6.4%
Industry	Finance	Banking, finance, insurance	4.0%
Industry	Real Estate	Real Estate	13.2%
Industry	Public	Public administration	5.0%
Industry	Education	Education	13.5%
Industry	Health	Health care	10.9%
Industry	Community	Community services	3.8%
Industry	Domestic	Domestic services	2.3%
Industry	Unknown industry	Industry unknown	2.9%
Occupation	Academic	Academic	13.6%
Occupation	Technical	Technical	14.7%
Occupation	Managerial	Managerial	5.7%
Occupation	Clerical	Clerical	18.6%
Occupation	Sales	Sales	20.5%
Occupation	Agricultural	Professional agricultural	0.6%
Occupation	Professional	Professional manufacturing	15.9%
Occupation	Unskilled	Nonprofessional	7.5%
Occupation	Unknown occupation	Occupation unknown	2.7%

Table 2: Cluster analysis results – sample means by cluster

	2 clusters		3 clusters		
Variable	1	2	1	2	3
Hourly wage	75.70	30.33	122.96	43.55	29.95
Age	47.72	36.71	49.64	41.59	37.57
Schooling	16.75	12.39	17.12	16.10	11.55
Female	0.50	0.52	0.36	0.64	0.46
Sector					
Arab	0.04	0.14	0.03	0.06	0.16
Yeshiva	0.02	0.01	0.01	0.03	0.01
Ethnicity					
Ashkenazi	0.31	0.09	0.39	0.21	0.08
Sefaradi	0.22	0.31	0.25	0.19	0.34
Native Israeli	0.19	0.25	0.19	0.25	0.23
USSR	0.24	0.21	0.14	0.28	0.19
Location					
Very peripheral	0.03	0.04	0.02	0.03	0.05
Peripheral	0.06	0.09	0.05	0.07	0.10
Intermediate	0.27	0.31	0.24	0.29	0.32
Central	0.15	0.14	0.17	0.13	0.14
Very central	0.49	0.42	0.51	0.48	0.40
Industry					
Agriculture	0.00	0.01	0.01	0.00	0.02
Manufacturing	0.14	0.16	0.18	0.12	0.18
Utilities	0.01	0.00	0.02	0.01	0.01
Construction	0.02	0.06	0.02	0.02	0.07
Commerce	0.06	0.16	0.07	0.07	0.17
Hospitality	0.01	0.07	0.00	0.02	0.08
Transportation	0.04	0.08	0.04	0.04	0.09
Finance	0.06	0.03	0.07	0.05	0.03
Real Estate	0.17	0.12	0.21	0.17	0.10
Public	0.06	0.04	0.07	0.05	0.04
Education	0.22	0.10	0.18	0.24	0.06
Health	0.14	0.10	0.10	0.16	0.08
Community	0.04	0.04	0.03	0.04	0.04
Domestic	0.01	0.03	0.00	0.01	0.04
Occupation					
Academic	0.35	0.03	0.46	0.23	0.01
Technical	0.24	0.11	0.15	0.29	0.05
Managerial	0.12	0.03	0.20	0.06	0.03
Clerical	0.13	0.22	0.10	0.19	0.21
Sales	0.08	0.28	0.04	0.12	0.31
Agricultural	0.00	0.01	0.00	0.00	0.01
Professional	0.06	0.22	0.04	0.06	0.26
Unskilled	0.02	0.11	0.01	0.03	0.13
Cluster size	5,551	10,739	1,895	6,008	8,387

Note: 607 observations with unknown industry or occupation were dropped.

Table 3: OLS regression results - dependent variable is ln(hourly wage)

Variable	1	2	3	4	5	6	7
Constant	1.122***	1.597***	1.126***	1.170***	1.068***	1.041***	1.542***
	(0.0560)	(0.0595)	(0.0560)	(0.0634)	(0.0557)	(0.0630)	(0.0658)
Age	0.066***	0.056***	0.066***	0.0662***	0.066***	0.066***	0.057***
	(0.0027)	(0.0025)	(0.0027)	(0.0027)	(0.0026)	(0.0026)	(0.0025)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female	-0.189***	-0.163***		-0.296***	-0.195***		
	(0.0099)	(0.0108)		(0.0523)	(0.0098)		
Arab	-0.152***	-0.088***		-0.156***	-0.181***		
	(0.0134)	(0.0141)		(0.0138)	(0.0585)		
Haredi	-0.321***	-0.275***		-0.313***	0.956***		
	(0.0433)	(0.0394)		(0.0420)	(0.0940)		
Female*Majority			-0.209***			-0.148***	-0.141***
			(0.0106)			(0.0520)	(0.0482)
Female*Arab			-0.071***			-0.459***	-0.298***
			(0.0237)			(0.1150)	(0.1080)
Male*Arab			-0.190***			-0.015	-0.022
			(0.0159)			(0.0720)	(0.0630)
Female*Haredi			-0.020			0.079	0.058
			(0.0450)			(0.2650)	(0.2740)
Male*Haredi			-0.512***			0.860***	0.309**
			(0.0618)			(0.1220)	(0.1240)
Schooling	0.077***	0.035***	0.077***				
	(0.0020)	(0.0020)	(0.0020)				
Schooling*Male				0.073***			
				(0.0031)			
Schooling*Female				0.081***			
				(0.0023)			
Schooling*Majority					0.081***		
					(0.0019)		
Schooling*Arab					0.003		
					(0.0046)		
Schooling*Haredi					-0.077***		
					(0.0050)		
Schooling*Male*Majority						0.083***	0.040***
						(0.0029)	(0.0028)
Schooling*Female*Majority						0.079***	0.038***
						(0.0024)	(0.0024)
Schooling*Male*Arab						-0.014**	-0.007
						(0.0059)	(0.0052)
Schooling*Female*Arab						0.029***	0.018**
						(0.0081)	(0.0076)
Schooling*Male*Haredi						-0.078***	-0.041***
						(0.0059)	(0.0057)
Schooling*Female*Haredi						-0.007	-0.008
						(0.017)	(0.0177)
Periphery/occupation/industry	no	yes	no	no	no	no	yes
Observations	16,290	16,290	16,290	16,290	16,290	16,290	16,290
R-squared	0.266	0.369	0.269	0.267	0.274	0.276	0.372

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Mixture regression results (model 7) - dependent variable is ln(hourly wage)

	Two-segment		Three- segment		
Variable	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
Constant	-0.143	1.991***	-0.289	2.466***	1.885***
	(0.3163)	(0.0637)	(0.3150)	(0.1320)	(0.1610)
Age	0.100***	0.042***	0.106***	0.028***	0.043***
	(0.0120)	(0.0025)	(0.0125)	(0.0043)	(0.0059)
Age squared/100	-0.090***	-0.039***	-0.101***	-0.029***	-0.034***
	(0.0144)	(0.0031)	(0.0152)	(0.0043)	(0.0065)
Female*Majority	-0.303	-0.106**	-0.057	0.006	-0.303***
	(0.2417)	(0.0482)	(0.2380)	(0.0629)	(0.0761)
Female*Arab	-0.168	-0.486***	0.181	-0.645***	-0.301
	(0.6572)	(0.1091)	(0.6800)	(0.2340)	(0.4590)
Male*Arab	0.380	-0.092	0.558*	-0.068	-0.288***
	(0.3008)	(0.0630)	(0.2900)	(0.0960)	(0.1030)
Schooling*Male*Majority	0.043***	0.040***	0.044***	0.018***	0.054***
	(0.0120)	(0.0031)	(0.0123)	(0.0046)	(0.0041)
Schooling*Female*Majority	0.052***	0.035***	0.038***	0.010***	0.059***
	(0.0132)	(0.0025)	(0.0129)	(0.0023)	(0.0048)
Schooling*Male*Arab	0.013	0.035***	-0.043*	-0.003	0.008
	(0.0221)	(0.0048)	(0.0230)	(0.0080)	(0.0080)
Schooling*Female*Arab	0.054	0.054***	-0.001	0.045***	0.010
	(0.0448)	(0.0074)	(0.0483)	(0.0163)	(0.0301)
Central region	0.125*	0.002	0.138**	0.003	-0.018
	(0.0639)	(0.0129)	(0.0642)	(0.0188)	(0.0183)
Intermediate region	-0.059	-0.059***	-0.052	-0.026*	-0.078***
	(0.0537)	(0.0105)	(0.0541)	(0.0156)	(0.0172)
Peripheral region	-0.024	-0.061***	-0.000	-0.053**	-0.065**
	(0.0767)	(0.0158)	(0.0753)	(0.0249)	(0.0257)
Very peripheral region	0.052	-0.074***	0.062	-0.077**	-0.069*
	(0.1262)	(0.0224)	(0.1280)	(0.0300)	(0.0378)
Academic occupation	0.957***	0.625***	0.940***	0.512***	0.601***
	(0.1191)	(0.0243)	(0.1110)	(0.0388)	(0.0508)
Technical occupation	0.806***	0.470***	0.786***	0.326***	0.490***
	(0.1154)	(0.0214)	(0.1070)	(0.0372)	(0.0475)
Managerial occupation	0.997***	0.643***	0.897***	0.526***	0.652***
	(0.1401)	(0.0299)	(0.1390)	(0.0467)	(0.0513)
Clerical occupation	0.585***	0.222***	0.548***	0.192***	0.196***
	(0.1067)	(0.0177)	(0.0991)	(0.0269)	(0.0458)
Sales occupation	0.506***	0.062***	0.441***	0.002	0.104**
	(0.1126)	(0.0175)	(0.1030)	(0.0231)	(0.0476)
Agricultural occupation	0.303	0.154**	0.329	-0.006	0.225**
	(0.3090)	(0.0596)	(0.3290)	(0.0783)	(0.0917)
Professional occupation	0.377***	0.056***	0.359***	0.090***	0.014
	(0.1138)	(0.0189)	(0.1070)	(0.0256)	(0.0442)

Continued on next page

Table 4 (continued)

Variable	Two-segment Seg. 1	Two-segment Seg. 2	Three- segment Seg. 1	Three- segment Seg. 2	Three- segment Seg. 3
Agriculture	-0.420*	-0.065	-0.400*	0.0578	-0.171*
	(0.2177)	(0.0465)	(0.219)	(0.0610)	(0.0898)
Utilities	0.345**	0.378***	0.380**	0.369**	0.368***
	(0.1742)	(0.0630)	(0.171)	(0.151)	(0.0742)
Construction	0.067	-0.023	0.110	0.0141	-0.0537*
	(0.1091)	(0.0184)	(0.112)	(0.0241)	(0.0307)
Commerce	0.047	-0.068***	0.0827	-0.0152	-0.108***
	(0.0890)	(0.0171)	(0.0911)	(0.0253)	(0.0259)
Hospitality	-0.262**	-0.087***	-0.176	-0.0384	-0.123**
	(0.1328)	(0.0239)	(0.128)	(0.0393)	(0.0506)
Transportation	0.185*	-0.007	0.192*	-0.0275	0.00708
	(0.1029)	(0.0194)	(0.106)	(0.0268)	(0.0296)
Finance	0.475***	0.168***	0.488***	0.167***	0.167***
	(0.1196)	(0.0267)	(0.127)	(0.0377)	(0.0376)
Real Estate	-0.040	-0.032	-0.0111	-0.0564**	-0.00445
	(0.0921)	(0.0194)	(0.0933)	(0.0251)	(0.0250)
Public	0.158	0.148***	0.243**	0.222***	0.0787**
	(0.1098)	(0.0234)	(0.113)	(0.0523)	(0.0356)
Education	-0.256***	-0.051**	-0.177*	0.0839**	-0.162***
	(0.0940)	(0.0200)	(0.0953)	(0.0408)	(0.0373)
Health	-0.293***	-0.134***	-0.275***	-0.0442	-0.178***
	(0.1010)	(0.0197)	(0.101)	(0.0291)	(0.0297)
Community	-0.049	-0.059**	-0.0316	0.00102	-0.0774*
	(0.1279)	(0.0273)	(0.127)	(0.0404)	(0.0408)
Domestic	-0.277	-0.226***	-0.0918	0.151***	-0.616***
	(0.2074)	(0.0344)	(0.201)	(0.0506)	(0.0673)
Sigma	0.928	0.376	0.910	0.244	0.355
Probability weight	0.177	0. 823	0.172	0.342	0.486
Predicted log-wage	3.387	3.688	3.269	3.464	3.764
AIC	3364971		3254460		
BIC	3365516		3255282		
Log Likelihood	-1682414		-1627123		
Observations	16,004		16,004		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Segment-specific means based on assigned segment

	Two-segment		Three- segment		
Segment	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
Hourly Wage	57.99	44.85	39.77	30.73	58.43
Age	39.90	40.54	37.44	41.39	40.32
Schooling	14.32	13.78	13.82	13.62	13.98
Female	0.488	0.515	0.543	0.503	0.516
Arab	0.049	0.117	0.062	0.117	0.116
Very Central	0.492	0.434	0.470	0.433	0.438
Central	0.128	0.143	0.132	0.139	0.146
Intermediate	0.281	0.300	0.293	0.302	0.296
Peripheral	0.061	0.086	0.067	0.086	0.085
Very Peripheral	0.038	0.037	0.039	0.039	0.035
Agriculture	0.012	0.011	0.011	0.013	0.009
Manufacturing	0.149	0.158	0.133	0.161	0.157
Utilities	0.006	0.007	0.006	0.007	0.008
Construction	0.029	0.051	0.032	0.053	0.049
Commerce	0.131	0.124	0.132	0.122	0.126
Hospitality	0.063	0.047	0.072	0.049	0.043
Transportation	0.055	0.068	0.048	0.070	0.067
Finance	0.050	0.041	0.041	0.038	0.045
Real Estate	0.160	0.135	0.142	0.140	0.134
Public	0.033	0.050	0.038	0.046	0.053
Education	0.145	0.133	0.166	0.115	0.143
Health	0.101	0.114	0.110	0.122	0.106
Community	0.047	0.038	0.049	0.040	0.036
Domestic	0.019	0.024	0.019	0.024	0.024
Academic	0.165	0.136	0.148	0.122	0.149
Technical	0.153	0.148	0.156	0.138	0.154
Managerial	0.071	0.059	0.049	0.054	0.066
Clerical	0.165	0.194	0.174	0.184	0.200
Sales	0.229	0.211	0.241	0.213	0.206
Agricultural	0.006	0.006	0.005	0.007	0.006
Professional	0.123	0.169	0.126	0.182	0.159
Unskilled	0.088	0.079	0.102	0.099	0.061
Sample share	0.081	0.919	0.087	0.393	0.520

Table 6: Segment-specific means based on weighted segments

	Two-segment		Three- segment		
Segment	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
Hourly Wage	52.00	44.60	46.02	32.92	54.99
Age	40.94	40.39	40.07	40.52	40.61
Schooling	14.14	13.76	13.98	13.61	13.92
Female	0.504	0.515	0.520	0.512	0.511
Arab	0.080	0.118	0.084	0.126	0.111
Very Central	0.460	0.434	0.454	0.427	0.441
Central	0.141	0.143	0.141	0.138	0.146
Intermediate	0.288	0.301	0.292	0.308	0.294
Peripheral	0.074	0.086	0.076	0.089	0.083
Very Peripheral	0.037	0.037	0.038	0.038	0.036
Agriculture	0.010	0.011	0.010	0.011	0.011
Manufacturing	0.156	0.157	0.148	0.159	0.159
Utilities	0.007	0.007	0.007	0.006	0.007
Construction	0.039	0.051	0.040	0.055	0.048
Commerce	0.125	0.125	0.128	0.127	0.122
Hospitality	0.048	0.048	0.053	0.050	0.044
Transportation	0.063	0.068	0.062	0.068	0.068
Finance	0.046	0.041	0.044	0.038	0.044
Real Estate	0.149	0.135	0.142	0.136	0.136
Public	0.045	0.050	0.047	0.047	0.051
Education	0.137	0.133	0.144	0.123	0.138
Health	0.111	0.113	0.111	0.117	0.111
Community	0.043	0.038	0.043	0.039	0.037
Domestic	0.021	0.024	0.022	0.025	0.024
Academic	0.156	0.134	0.149	0.121	0.146
Technical	0.156	0.146	0.155	0.135	0.155
Managerial	0.068	0.058	0.060	0.053	0.064
Clerical	0.180	0.194	0.183	0.190	0.196
Sales	0.214	0.211	0.221	0.218	0.204
Agricultural	0.005	0.006	0.005	0.006	0.006
Professional	0.143	0.170	0.145	0.181	0.161
Unskilled	0.077	0.080	0.082	0.095	0.068
Observations	16004	16004	16004	16004	16004