Journal of Comparative Economics **30**, 276–295 (2002)
doi:10.1006/jcec.2002.1774

The Gender Wage Gap in Bulgaria: A Semiparametric Estimation of Discrimination[1]

Dean Jolliffe

Economic Research Service, U.S. Department of Agriculture, Room S-2059, 1800 M Street NW, Washington, DC 20036; and William Davidson Institute, University of Michigan, 724 East University Avenue, Wyly Hall, Ann Arbor, Michigan 48109-1234
E-mail: Jolliffe@ers.usda.gov

Received August 25, 2001; revised March 6, 2002

Jolliffe, Dean—The Gender Wage Gap in Bulgaria: A Semiparametric Estimation of Discrimination

Bulgaria's transition to a market economy has coincided with a large increase in wage inequality. This increase may be due to managers rewarding more productive workers or it may be the result of rewarding noneconomic characteristics such as sex. Data from the 1995 Bulgaria Integrated Household Survey reject the hypothesis of no sex discrimination. Using separate wage regression estimates for men and women, an Oaxaca decomposition indicates that men's wages are 24% higher than women's wages and 86 to 105% of this differential is due to differences in how men and women are rewarded for the same characteristics. *J. Comp. Econ.*, June 2002, **30**(2), pp. 276–295. Economic Research Service, U.S. Department of Agriculture, Room S-2059, 1800 M Street NW, Washington, DC 20036; and William Davidson Institute, University of Michigan, 724 East University Avenue, Wyly Hall, Ann Arbor, Michigan 48109-1234. Published by Elsevier Science (USA)

Key Words: human capital; Bulgaria; wage differentials; discrimination; gender; sample selection model; type 3 tobit; semiparametric estimation.

Journal of Economic Literature Classification Numbers: I2, J16, C14.

1. INTRODUCTION

The transition in Central and Eastern Europe from centrally planned to market economies has resulted in large changes in the distribution of wages and increases

[1] I thank Peter Orazem, Anna Ratcheva, Nauro Campos, Josh Winicki, Craig Gundersen, Linda Ghelfi, Leslie Whitener, John Bonin, and two anonymous referees for comments. I am grateful to Avgoustina Tchoutkina for providing me with the 1998 Bulgarian Statistical Yearbook, and I also thank Bohdan Krushelnytskyy and Anastassia Semykina for excellent research assistance. The views and opinions expressed in this paper do not necessarily reflect the views of the Economic Research Service of the U.S. Department of Agriculture.

0147-5967
Published by Elsevier Science (USA)

in inequality. Milanovic (1998) states that the Gini coefficient of disposable income increased an average of 9 points, from 24 to 33, over the period from 1987–1988 to 1993–1995.[2] In the case of Bulgaria, Milanovic (1999, Table 8) reports that the income Gini coefficient increased 10 points, from 21.7 to 31.7 between 1989 and 1995. By decomposing this change in the Gini coefficient, he demonstrates that 78% of the increase in inequality is due to the change in the distribution of wages.

Jones (1991, Table 5) provides evidence of the dramatic increase in wage inequality in Bulgaria by examining the coefficient of wage variation for 11 employment sectors. His measure of wage dispersion is 0.12 for 1960, 0.12 for 1970, 0.10 for 1980, and 0.12 for 1987.[3] In the early transition years of 1989 and 1990, Jones shows that this measure increased to 0.31 and 0.26, respectively. Beleva et al. (1995, Table 5-5) examine the coefficient of wage variation across 18 industries and report that this measure increased by 71% from 1990 to 1992.

In addition to the changing distribution of wages, the wage determination process in transitional countries has also been changing by becoming increasingly more decentralized. Svejnar (1999) provides an overview of labor markets in Central and East European (CEE) countries during transition. In the case of Bulgaria, Brainerd (2000) notes that a relatively decentralized wage-setting system has evolved in Bulgaria despite national agreements to establish wages floors and ceilings. Jones (1991) asserts that the large increase in the dispersion of wages in 1989 was due in part to new legal arrangements, i.e., the New System of Basic Wages, that gave managers much more control over determining the wages of their employees.

The increased freedom of managers to determine wages, along with the rise in wage inequality, heightens the need to monitor issues related to social equity in labor markets. The major concern of this paper is sex-based wage discrimination. The importance of this issue is enhanced as discussions continue on Bulgaria's progress towards accession to the European Union (EU). The ability to join the EU is contingent on Bulgaria's ability to institute a legal and institutional framework, referred to as the *acquis communitaire*, that is in accord with the objectives of the EU. An important aspect of the *acquis communitaire* is the section on "Social Policy and Employment" that requires equal treatment for women and men in labor markets. Bulgaria is viewed as having made very little progress in this area and must undertake major efforts to achieve compliance. The Commission of the European Communities (2001) provides more details on Bulgaria's progress towards accession. In particular, Chapter 13 of this report contains an extensive discussion of social policy and employment in Bulgaria.

The purpose of this paper is to estimate a model of wage determination in Bulgaria and test the null hypothesis that the pay structure for men and women is the same. A rejection of this null hypothesis is evidence of sex discrimination

[2] Milanovic notes that the rate of this increase was three times as great as the increase in inequality experienced in the West during the 1980's.

[3] Jones asserts that these figures show a tight compression of wages relative to Western Europe, and also relative to other East European countries.

because men and women with the same levels of education, experience, and other economically relevant factors receive different pay. After testing for discrimination, the observed male–female wage gap is decomposed into two components. The first results from male–female differences in economic characteristics, such as education. The second results from different pay structures, or discrimination. This decomposition provides evidence as to whether wage inequality is caused only by managers rewarding more productive workers, and thus improving economic efficiency, or whether wage inequality is partially dependent on managers rewarding noneconomic characteristics and thereby hampering the transition.

Although there has been a growing literature on gender wage differentials in transition countries, relatively less work exists for Bulgaria due to data limitations. Brainerd (2000) reports that in 1992 postreform Bulgaria, the female to male wage ratio at the means was 78.4% but she does not estimate how much of this difference in wages can be explained by differences in human capital characteristics. Brainerd shows that this estimate is similar to some of the other postreform CEE countries. For example, the gender wage ratio for Hungary in 1991 was 75.1%; for Poland in 1992, it was 81.2%.

Using 1994 data from Romania, Paternostro and Sahn (1999) find that women working in urban areas earn on average 80% of what men earn. An Oaxaca decomposition of their estimates show that only 21% of this wage gap can be explained by relevant earnings characteristics, such as education and experience.[4] Jurajda (2001) shows that, in 1998, Czech women working in the nonpublic sector earn on average about 70% of what men earn. Only 36% of this difference can be explained by difference across the sexes in human capital attainment as well as differences in occupation types and firm characteristics, such as size and ownership status. Jurajda demonstrates that women working in the nonpublic sector in Slovakia fare somewhat better as they earn on average about 77% of the mean male wage. While the wage gap is smaller in Slovakia than in the Czech Republic, the proportion of this gap that can be explained by economic characteristics of the worker, occupation classification, and firm attributes is about the same at 38%.

Using data from 1993 for four CEE countries, Pailhé (2000) shows that differences in observed characteristics of male and female workers does not explain much of the difference in wages. Pailhé reports that the difference in the mean of log wages is 0.23 for Hungary and 0.30 for Czech Republic. Differences in education and experience across the sexes explain 26% of the gap in Hungary and 19% in the Czech Republic. The log wage gaps are similar for Poland at 0.30 and Slovakia at 0.29, but education and experience explain even less of the gap for these countries. In the case of Slovakia, human capital differences explain only

[4] An important methodological difference between their work and this paper is that Paternostro and Sahn estimate wages using the Heckman selection model with maximum likelihood. This estimator is inconsistent in the presence of heteroscedasticity.

8% of the gap and Polish female wage earners would actually have higher wages if remunerated similarly for education and experience.[5]

For Russia in 1996, Reilly (1999) finds that the difference in the average log monthly wage between men and women was very large at 0.36. An Oaxaca decomposition of this difference shows that only 12% of this gap can be explained by sex differences in schooling, age proxying for experience, and occupation type. Reilly shows that the results are quite similar for Russia in 1992, suggesting that there was little change in discrimination during the transition period. Similarly, Linz (1996) analyzes 1992 data from Taganrog, Russia, and finds that the differences in male and female wages cannot be explained by human capital characteristics.

These results contrast with evidence from the United States for which Blau and Kahn (2000) report that the wage differential between men and women was 27.6% in 1988.[6] Their decomposition of this gap suggests that one-third of it can be explained by differences across the sexes in human capital attainment. In particular, they find that the primary factor was differences in experience levels with men having on average 4.6 years more experience. When the specification is expanded to include information on occupation, industry, and union status, the model explains over 60% of the wage gap. Similarly, Sicilian and Grossberg (2001) find that approximately 60% of the U.S. gender wage gap can be explained by differences in human capital attainment and market characteristics. They demonstrate that male–female differences in human capital characteristics are the most important factors in explaining the wage gap.

The discrimination literature suggests that while the gender wage gaps in CEE countries are similar to those found in the United States and the rest of Europe, somewhat less of the CEE wage differentials can be explained by human capital characteristics. This difference suggests that discrimination might be a more important determinant in explaining the wage gap in CEE countries than in the U.S. and Western Europe. Sziráczki and Windell (1992) provide qualitative evidence that women in Bulgaria may face discriminatory practices. They report results from a survey of managers who were asked whether they had a preference for men or women when hiring for production or professional work.[7] While 25% stated a preference for hiring women for skilled production work, 54% stated a preference for hiring men. In recruiting professional staff, 59% of the managers stated that they would prefer to hire a man rather than a woman, while only 3% stated a preference for hiring a woman.

[5] Pailhé also estimates a model that includes professional and sectoral variables; this model explains much more of the wage gap. For example, in the case of Poland, the model explains 53% of the log wage differential. This result suggests that women with equal skills sort into lower paying sectors and occupations.

[6] Blau and Kahn also show that the female to male wage ratio increased to 0.76 in the U.S. by 1996. This compares to an average value of 0.77 for 16 major developed countries between 1994 and 1998.

[7] Completed in 1992, the Bulgarian Labor Flexibility Survey was based on 500 enterprises covering the manufacturing sector in 4 of the 9 national districts, including Sofia.

The plan of this paper is as follows. Section 2 reviews the characteristics of the 1995 Bulgaria Integrated Household Survey (BIHS), which is the data set used in this paper to test for sex discrimination. Section 3 discusses the methodology used to estimate wage functions. Wage determination is modeled with a correction for sample selection as a Type III Tobit and estimated with the Honoré et al. (1997) semiparametric estimator. Unlike the classic Heckman correction for sample selection, this estimator is consistent in the presence of heteroscedasticity. Section 4 reports the main results demonstrating that the difference in the mean of log wages for men and women is 0.22 and that 86 to 105% of this difference is due to discrimination and not difference in economic performance. Because this analysis is cross-sectional, the increase in wage inequality cannot be attributed to an increase in sex discrimination. However, the results do show that the presence of sex discrimination is an important determinant of the level of wage inequality. Section 5 provides a brief conclusion.

2. DATA

The data used in this paper are from the nationally representative Bulgaria Integrated Household Survey (BIHS) and were collected during the summer of 1995. The Gallup Organization in Sofia managed the survey and the World Bank provided technical assistance. The sample frame, from which the sample was drawn, is based on the 1992 general census of Bulgaria. The planned sample size was 2500 households, which were drawn randomly in a two-stage process. Not all households agreed to participate or could be located so that the actual sample size was 2466 households with 7199 individuals. The midyear population count for 1995 was 8,272,339 persons (U.S. Census Bureau, Table 001, 2001), which means that the raising factor, i.e., the ratio of the population to the sample size, is 1149.

To focus on working age adults, 2491 persons who are less than 17 years of age or greater than 65 years of age are dropped from the sample.[8] The lower bound of 17 years is chosen since school is compulsory through the age of 16. Further, in the BIHS sample, there are no wage earners under 17 years of age who have completed their schooling. The upper age bound of 65 is higher than the official age of retirement, but the BIHS data indicate that many individuals continue to work for a wage after the age of retirement. Nonetheless, by excluding persons over 65 years of age, less than 0.5% of the sample of wage earners is lost.

Two hundred forty-four continuing students are also dropped from the sample so that only working-age persons who have completed schooling remain in the sample. Finally, there are 215 persons, which is less than 5% of the sample, that are missing wage, education, or household composition data; these observations are also excluded from the analysis. The resulting, effective sample size is

[8] Narrowing the sample on age will not introduce self-selection bias as age is exogenous to the individual.

4249 persons of which 44% (1874) are wage earners.[9] In terms of the sex composition of the sample, 2182 (51%) of the 4249 persons are female, and 892 (48%) of the 1874 wage earners are female. Sixty-nine percent of the effective sample, and 78% of the wage earners, live in urban areas.

Using the BIHS data, gross wage income is constructed as the sum of wage payments before taxes, wage adjustments for children, allowances for transportation costs, and other payments. To this sum, imputed values for paid leave and subsidized vacations are added. The resulting estimate of monthly, average wage income for the 1874 wage earners is 6822 Leva. Using the 6984 observations without missing data, this implies that yearly, per capita wage income is 21,966 Leva. This estimate compares closely with data from the National Statistical Institute (1998, Table V-1) report of per capita, yearly wage income at 22,243 Leva in 1995.[10] This is a difference of less than 1.3%, which is well within the sampling error of the surveys. The estimate of wages is gross wage income divided by hours of wage work.

Relative to 1996 and 1997, inflation was quite low in 1995 at an annual rate of 33%. For the 3-month span of the survey, the consumer price index rose by 2% (National Statistical Institute, 1998, Table VI-1). To correct for this, all reported prices have been deflated to the first month of the survey, May 1995. This relatively small adjustment for inflation has little effect on the reported estimates.

More important than the temporal change in prices is the spatial differences in price levels. The National Statistical Institute does not report a spatial-price index, although the cost of living is much higher in Sofia than elsewhere and urban prices are higher than rural prices. The World Bank (1996) asserts that the cost of living in Sofia is 23% greater than in the region of Haskovo and that urban prices are 6% higher than rural prices on average. Unfortunately the methodology used to construct these estimates is not documented so that there is no commonly accepted index for spatial price adjustment. Instead of using an undocumented index of spatial price variation, variation in prices is controlled for with dummy variables for the political regions of Bulgaria as well as for urban areas.[11] In addition to controlling for spatial-price variation, these regional dummies will control for any region-specific, fixed effect. Two examples of potentially important regional effects relate to the capital intensity of production and school quality. Both of these factors affect wages and both are likely to vary across regions.

[9] Wage earners are defined as all individuals who are working for a salary or commission for somebody else. This excludes all self-employment activities.

[10] The estimate of yearly wage income from the National Statistical Institute is very similar in construction to the estimate used in this paper. It is an aggregate measure of cash and in kind payments and includes a value for annual vacations, public holidays, paid leave, and overtime, but does not include tax payments.

[11] The BIHS sample covers nine political regions including Sofia City, Bourgas, Varna, Lovech, Montana, Plodviv, Russe, Sofia Region, and Haskovo.

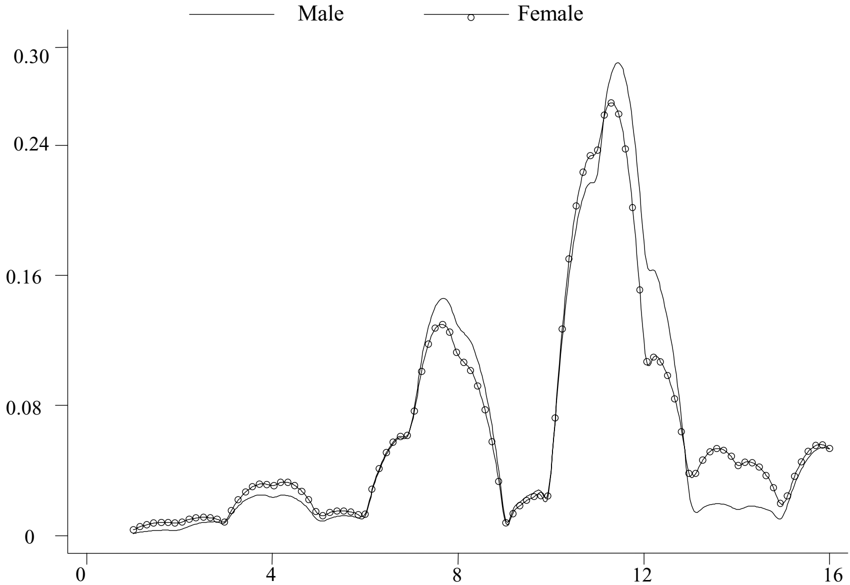

FIG. 1. Years of schooling by sex, kernel estimate of the density function. Tails of the distribution are trimmed at the 1st and 99th percentile resulting in a sample size of 2088 women and 1982 men. The Epanechnikov kernal is used with a bandwidth of 0.46. This width is 80% of the optimal width for estimating the density of a normally distributed variable with the Gaussian kernal.

Education levels in Bulgaria have historically been quite high with enrollment rates for primary and middle schools on par with, or better than, West European countries. In addition to successful enrollment rates, the Bulgarian schooling system has also been characterized as providing girls access to the same opportunities as boys. Laporte and Ringold (1997) provide more details on access and enrollment rates by sex. The BIHS data indicate similar access to schooling opportunities with education outcomes that are quite similar across sex. In 1995, the BIHS data show that the average level of schooling attained is 10.48 years for males and 10.40 years for females. The median (11 years) and interquartile range (8, 12 years) of schooling attainment are the same across the sexes. As further evidence of equality of schooling outcomes, Fig. 1 provides kernal density estimates of the male and female distributions of school attainment. Over the entire range of years of schooling, the distributions are very similar.

To specify educational attainment in the wage regression, it is useful to examine the Bulgarian schooling system. Education is compulsory for children up to the age of 16. Primary school begins at the age of 6 or 7 and runs for 4 years. Middle school follows for 4 more years and, at the end of grade eight, students are tested in Bulgarian and mathematics.[12] The results of these tests determine whether

[12] In spite of the compulsory attendance requirements, the BIHS data show that 15% of the population between the ages of 17 and 65 have less than 8 years of schooling.

the students progress on to vocational, technical, or more general academic, secondary school.[13] This multitrack system is similar to that used in Germany and much of Europe, including other CEE countries.[14] Throughout the first half of the 1990's, approximately 40% of the enrolled secondary students attended general academic schools, while about 30% attended technical schools and the remaining 30% attended vocational schools (Laporte and Ringold, 1997, Table 1).

To control for the various types of secondary schooling, dummy variables are included in the wage regression model to designate if the person attended a vocational, technical, or general secondary school. In many countries,[15] the returns to primary schooling differ significantly from secondary schooling so that it is desirable to specify a functional form that allows for this possibility. Years of primary and middle school are combined as one variable and the years of secondary and postsecondary schooling are summed to create a separate regressor. This break allows for some nonlinearity in the returns to years of schooling and the break point appears reasonable given the institutional details of the Bulgarian schooling system.

In summary, the nationally representative BIHS data allow for careful measurement of wages that captures important supplemental benefits and matches estimates from independent data sources. Through the use of regional dummy variables, the estimation controls for region-specific fixed effects such as spatial price variation. The BIHS data also include information on sector of employment and this information is used to create 10 dummy variables for each major sector, including agriculture, manufacturing, construction, and science. Economic returns and the proportion of men and women vary by sector so it is useful to control for employment sectors when estimating discrimination. Finally, the choice of creating two variables for pre- and postsecondary schooling allows for some flexibility in the functional form of the wage equations.

3. ESTIMATION

Wage equations are estimated using a standard Mincer equation, taking the form

$$\ln(w_i) = \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 X_i + \varepsilon_i, \tag{1}$$

[13] The postmiddle school alternatives are actually a bit more complex. Secondary school options include *gymnasium*, which is an academic secondary school; specialist schools, which are also academically oriented but are more specialized and selective; technical schools, which have similar curriculum to *gymnasium* plus an additional year of vocational training; and finally vocational schools, which specialize in industrial skills and craft training. For more details, see World Bank (2000).

[14] Filer et al. (1999) provide a description of similarly multitracked secondary school systems for the Czech and Slovak Republics. Notable exceptions to this type of school system are the single-track systems of the United States and much of the former Soviet Union.

[15] Psacharopoulos (1994) reports returns to education by primary and secondary schooling for 78 countries.

where the i subscript denotes the individual; w is wages; S is a vector of two variables, one for years of presecondary and another for postsecondary schooling; E is experience; and X is a vector of control variables including dummies for the 3 types of secondary schooling, the 10 sectors of employment, the 9 geographic regions, and a dummy to indicate urban areas.

The sample contains 4249 working-age individuals of which 1874 are wage earners. As is well known, OLS estimation of Eq. (1) on a sample of wage earners is likely to result in biased estimates due to sample selection. The sample selection problem is typically modeled as

$$w^* = x'\beta + \varepsilon \tag{2}$$

$$y = z'\gamma + \mu, \tag{3}$$

where w^* is the unobserved wage offer and y is a variable that indicates whether the individual is a wage earner.

In the Type II Tobit model, y is a dummy variable; for the Type III Tobit model, y is a continuous variable that is censored at zero with only positive outcomes observable. In this paper, y is the log of hours worked in wage employment implying that wage determination is modeled as a Type III Tobit model. Hours of labor force participation indicate whether the individual selects into wage employment and also the intensity with which a person is engaged in this activity. The BIHS data indicate that 64% of the wage laborers work 160 hours per month; the range is between 32 and 280 hours per month.

While w^* is the variable of interest, we observe only wages for actual wage earners, w, i.e.,

$$\begin{aligned} w &= w^* && \text{if } z'\gamma + \mu > 0 \\ w &= 0 && \text{otherwise.} \end{aligned} \tag{4}$$

When estimating wages on the sample of wage earners, the conditional expectation of w^* is

$$E(w^* \mid x, z, y > 0) = x'\beta + E(\varepsilon \mid x, z, \mu > -z'\gamma) \tag{5}$$

and the zero-mean restriction placed on the error term, ε, will not hold in general. Heckman (1974) proposes a way to correct for this sample-selection bias and to restore the zero-mean property of the residual by adding a selection-bias-correction term, or Mill's ratio, to the regression. A problem with this approach, as discussed by Hurd (1979) and Nelson (1981), is that the Heckman estimator, along with the Tobit estimator, are biased when the assumption of homoscedastic errors is violated.

In the case of the BIHS data, as with almost any multistage design survey, the assumption of homoscedastic residuals will rarely be tenable. The clustering of observations from the multistage aspect of the sample design results in observations that are more similar within clusters than across clusters. Deaton (1997, p. 79) asserts, "It is a fact that regression functions estimated from survey data are typically not homoscedastic." Indeed, the p-value from the Pagan–Vella test statistic (1989) estimating the wage regression rejects strongly the assumption of homoscedasticity ($p < 0.001$).[16]

Honoré et al. (1997) propose an estimator based on sample trimming for the Type III Tobit model that is robust to heteroscedasticity.[17] The estimator is similar conceptually to the Heckman two-step approach and employs the insights of Powell (1986) concerning trimmed estimators. In the first step, the selection equation, (3), is estimated via any estimator that provides consistent estimates for the censored regression model in the presence of heteroscedasticity. These authors propose using either Powell's censored least absolute deviations (CLAD) estimator (1984) or Powell's symmetrically trimmed least squares (STLS) estimator (1986).

The results from this estimation are used to trim from the sample all observations in which y is either equal to zero or greater than twice its predicted value, in other words, all observations where $0 < y < 2z'\hat{\gamma}$ are retained. The result of this trimming is that β is now estimated only for the observations where $-z\gamma < \mu < z\gamma$; this trimming restores the zero-mean condition on ε. The second stage of the Honoré et al. estimator is to estimate (2) on the trimmed sample using either ordinary least squares (OLS) or least absolute deviations (LAD).

For the first-stage estimates of the Honoré et al. estimator, the STLS estimator is used rather than the CLAD estimator because the CLAD estimator requires that the median be observed. In the case of the BIHS data, the median is censored since only 44% of the working-age sample are wage earners. Powell (1986) shows that minimizing the objective function, $S_T(\gamma)$, given by

$$S_T(\gamma) = \sum_{i=1}^{N}[y_i - \max(0.5y_i, z_i'\gamma)]^2 + \sum_{i=1}^{N} I(y_i > 2z_i'\gamma)[(0.5y_i)^2 - (\max(0, z_i'\gamma))^2] \tag{6}$$

yields implicitly a minimum to the normal equation and the $\hat{\gamma}$ which minimizes $S_T(\gamma)$ is the STLS estimator.

The results from the STLS estimation are informative about the selection decision into wage labor, i.e., Eq. (3). They also serve to trim the sample for the second step of the Honoré et al. estimator, which in this case is a Type III Tobit using symmetrically trimmed least squares, hereafter referred to as T3T-STLS.

[16] The Pagan–Vella test is a modification of the Breusch–Pagan test that can be used for cases like the sample-selection model in which the zero-mean assumption placed on residuals fails to hold.

[17] They make the general argument that the Type III Tobit model is intuitively more appealing than the Type II Tobit model because it exploits more information in the selection variable.

With the sample trimmed, the zero-mean condition of the residuals is restored and OLS estimation of the wage equation, (2), results in consistent point estimates.

However, the OLS estimate of the variance-covariance matrix will be inconsistent because it does not account for the first-stage trimming of the sample. To provide consistent variance estimates, bootstrap estimates are found by resampling the design matrix and following the two-step method to solve for the T3T-STLS estimates.[18] To denote the precision of the point estimates, two measures from the bootstrap method are provided. The first is the standard error of the empirical density function of the parameter estimate. For the second measure of precision, the parameter estimates are superscripted with *, **, or *** if the (two-tailed) 90th, 95th, or 99th percentile of the bias-corrected empirical density function excludes zero.

4. RESULTS

4.1. Selection Model and Wage Estimates

Identification of the selection model requires finding variables that explain the decision to work in the wage sector but are also not directly determinants of wages. Becker (1965) suggests that one factor influencing the reservation wage and thereby the decision to participate is nonemployment income.[19] The BIHS data have information on the net level of remittance income, income from real estate assets, and the amount of social assistance income in the household. In addition to these three variables, the model in Heckman (1974) on the selection of women into the wage market is followed. The number of young children of all ages up to 5 and the number of children older than 5 and less than 14 are also included in the specification.[20] The motivation for this inclusion is that younger children require care and older children have the potential to supply labor for household work.

Table 1 presents the STLS estimates of the selection model in which the dependent variable is the log of hours in wage work. This variable is censored at zero if the individual does not engage in wage employment and therefore has zero wages for the second-stage, wage regressions. As the purpose of this paper is to examine wage differentials by sex, the wage and selection models are estimated for men and women separately as well as for the combined sample. For all three models, the nonwage income variables and/or the family composition variables

[18] For a general discussion of the bootstrap method, see Efron and Tibshirani (1993).

[19] The decision to exclude from the model information on other sources of employment income is based on the assumption that the decision to work in the wage market is separable from the decision to work in other markets. This assumption is made to simplify the model and to help interpret the results. For more details on the assumption of separability in household labor models, see Singh et al. (1986).

[20] Additionally, there may be important cultural differences or attitudes towards work that vary by ethnicity so that dummy variables for Turkish and Romany persons could also be included in the model. This specification was tested and the dummy variables were statistically insignificant in affecting either the level of male or female hours of wage work.

TABLE 1
Log Hours in Wage Work, the Selection Model (Symmetrically Trimmed Least Squares)

	Full sample		Men		Women	
	STLS	(Std. error)	STLS	(Std. error)	STLS	(Std. error)
Years, lower school	0.030*	(0.0182)	−0.002	(0.0421)	0.037	(0.0478)
Years, secondary school	0.013	(0.0174)	0.039**	(0.0165)	−0.029	(0.0186)
Experience	0.240***	(0.0132)	0.039***	(0.0105)	0.018	(0.0130)
Experience squared	−0.003***	(0.0003)	−0.001***	(0.0003)	0.000	(0.0004)
Dummy: Female = 1	−0.988***	(0.0705)				
Dummy: Urban location	−0.261***	(0.0771)	−0.166*	(0.0880)	0.038	(0.1043)
Children 0 to 4	−0.168**	(0.0836)	0.097	(0.0723)	−0.840***	(0.0907)
Children 5 to 14	0.039	(0.0436)	−0.007	(0.0489)	0.084*	(0.0512)
Social benefit income	−0.001	(0.0166)	−0.034*	(0.0205)	0.042*	(0.0222)
Remittance received	0.002	(0.0240)	−0.039*	(0.0202)	−0.012	(0.0225)
Rents from real estate	0.007	(0.0095)	−0.003	(0.0119)	−0.083***	(0.0285)
Intercept	−0.028	(0.0269)	−0.934	(0.0338)	−1.431***	(0.0353)
Joint tests of controls [p-values]						
Three types of secondary school		[0.0000]		[0.3266]		[0.0069]
Eight political districts		[0.0000]		[0.0000]		[0.0000]
Nine sectors of employment		[0.0000]		[0.0000]		[0.0000]
Sample size/No. of Obs. > 0	4249/1874		2067/982		2182/892	
Adj. R^2 (for uncensored obs.)	0.49		0.80		0.90	

Note. (1) All nonwage income is in thousands of leva. (2) The dependent variable is the log of hours in wage employment, which serves as an indication of both whether the individual is engaged in wage labor and also the intensity. (3) The adjusted R^2 is calculated only for the positive observations (excludes censored observations). (4) Parameters are superscripted with *, **, or *** if the p-value is less than 0.1, 0.05, or 0.01.

are significant. In the case of the combined sample of men and women, the number of children in the family up to 4 years of age is negative and significant suggesting that an increased number of children decreases an adult's ability to engage in the wage sector. The results for men differ somewhat in that the family composition variables have no effect on their decision to participate in the wage sector, but increased levels on nonwage income for the family reduces the men's hours in wage labor. In particular, remittance and social benefit income both have negative effects of similar magnitude and both are significant.

In the case of women, young children have a large and significant negative effect on participation while mid-aged children have a positive effect. This finding suggests that women reduce their wage participation hours to tend to young children but that mid-aged children free up some of the time demands placed on women allowing them to work more in the wage sector. Women's participation in the wage sector is also affected by nonwage income. Rents from real estate have

TABLE 2
Wage Regression Model, Returns to Primary and Secondary Schooling (Type III Tobit with Symmetrically Trimmed Least Squares, T3T-STLS)

	Full sample		Men		Women	
	T3T-STLS	(Std. error)	T3T-STLS	(Std. error)	T3T-STLS	(Std. error)
Years, lower school	0.056***	(0.0273)	0.056*	(0.0249)	0.035	(0.0655)
Years, secondary school	0.060***	(0.0076)	0.049***	(0.0098)	0.081***	(0.0115)
Joint test of school [p-value]		[0.0000]		[0.0000]		[0.0000]
Experience	0.008	(0.0077)	0.020*	(0.0078)	0.030***	(0.0118)
Experience squared	0.000	(0.0002)	0.000*	(0.0002)	−0.001**	(0.0003)
Dummy: Female = 1	−0.183***	(0.0333)				
Dummy: Urban location	0.107***	(0.0358)	0.069	(0.0531)	0.113*	(0.0718)
Intercept	3.465***	(0.3172)	3.099***	(0.3362)	2.821***	(0.6204)
Joint tests of controls [p-values]						
Three types of secondary school		[0.0415]		[0.6558]		[0.7317]
Eight political districts		[0.0000]		[0.0000]		[0.0465]
Nine sectors of employment		[0.0002]		[0.0116]		[0.7793]
Obs. > 0/trimmed sample	1874/1072		982/912		892/873	
Adjusted R^2	0.22		0.17		0.25	

Note. (1) The dependent variable is the log of wages. (2) Parameters are superscripted with *, **, or *** if the two-sided 90th, 95th, or 99th bias-corrected, empirical density function excludes zero. (3) Standard errors and p-values are found by treating the bootstrap samples as an estimate of the empirical variance-covariance matrix. (4) The adjusted R^2 is over the trimmed sample.

a negative and significant effect although increased levels of social benefits have a counterintuitive positive effect.[21]

Table 2 presents the wage regression results with the sample selection correction through the trimming method. The first column presents the national estimates with a dummy variable for whether the individual is male or female. The point estimate on the female dummy suggests that women's wages are 18% lower than men's wages even after controlling for schooling, type of secondary school, experience, sector of employment, and region of residence. However, this estimate is a bit naive because it assumes that the only difference between the wage determinants is in the intercept and that all of the slope parameter estimates are the same.

To allow for the likely event that the difference in pay between men and women is affected by more than just a shift in the intercept, the second and third columns in Table 2 estimate sex-specific wage regressions. From these results it is clear

[21] This result is linked perhaps to the result that men work less with increased levels of social benefit income and that most women are in households with a male spouse who is about 3 years older. It is also possible that men are more likely to be disabled from work, which increases social benefits, but not by an amount sufficient to support the family so that women must work.

that there are important differences. The Hausman test statistic rejects strongly the assumption that parameter estimates for the female regression are equal to those from the male wage regression.[22] More specifically, the intercept for men is greater than for women, which is not too surprising given the negative point estimate on the female dummy in the pooled sample. In terms of experience and education, women receive greater returns, as a percent of their wages, from experience and secondary schooling. The rate of return to an additional year of secondary schooling for women is 8.1% in contrast to 4.9% for men.

The evidence of a heteroscedastic error structure found above implies that the standard Heckman selection estimates, either two-step or maximum likelihood, are inconsistent. Nonetheless, it is still reasonable to ask whether the consistent T3T-STLS differ significantly from the Heckman estimates. For both the male and female regression results, Hausman tests reject strongly the null hypothesis that the Heckman estimates are equal to the T3T-STLS estimates.[23] As an example, one of the more striking differences in the parameter estimates is in the returns to experience. The Heckman estimate is 85% greater than the T3T-STLS estimate for the male regression. As another example from the male regression, the Heckman estimate for the return to years of secondary schooling is 18% greater than the T3T-STLS estimate.

4.2. Oaxaca Decomposition

Given differences across male and female estimates for wage functions, Oaxaca (1973) proposes a method to decompose the observed gender wage gap into two components. The first results from differences in education and experience levels, and any other explanatory variable, of men and women. The second component consists of differences that are considered to be the result of current labor-market discrimination. However, interpreting the model residual as discrimination needs to be done with caution. If there is any omitted variable that has a positive effect on wages, and if men are more highly endowed with this characteristic, the results from the decomposition would overestimate discrimination. Alternatively, if some of the factors in the model are themselves affected by discrimination, the analysis could underestimate discrimination. For example, if women are more likely to be fired in economic downturns, or if they have less access to the types of schooling deemed more valuable by the market, the decomposition may underestimate discrimination.

The Oaxaca decomposition rests on the fact that the OLS regression line runs through the mean, and the difference in the mean of log wages between men and

[22] The test statistic is formed as $(\hat{\beta}_{\text{female}} - \hat{\beta}_{\text{male}})' (\hat{V}_{\text{female}} - \hat{V}_{\text{male}})^{-1} (\hat{\beta}_{\text{female}} - \hat{\beta}_{\text{male}})$ and the resulting p-value from testing for equality of $\hat{\beta}_{\text{female}}$ and $\hat{\beta}_{\text{male}}$ is equal to 4.3e-10.

[23] The Hausman test statistic for comparing the male regression results across the two estimators is 120 with 26 degrees of freedom and a p-value of 5.2e-14. The Hausman test statistic comparing the Heckman female estimates with the T3T-STLS female estimates is 80 and this has a p-value of 2.1e-7.

women can be written as

$$\begin{aligned}\overline{\ln(w_{\mathrm{m}})} - \overline{\ln(w_{\mathrm{f}})} &= (\bar{x}'_{\mathrm{m}} - \bar{x}'_{\mathrm{f}})\hat{\beta}_{\mathrm{m}} + \bar{x}'_{\mathrm{f}}(\hat{\beta}_{\mathrm{m}} - \hat{\beta}_{\mathrm{f}}) \\ 0.22 &= 0.03 + 0.19\end{aligned} \tag{7}$$

or, similarly this can be expressed as

$$\begin{aligned}\overline{\ln(w_{\mathrm{m}})} - \overline{\ln(w_{\mathrm{f}})} &= (\bar{x}'_{\mathrm{m}} - \bar{x}'_{\mathrm{f}})\hat{\beta}_{\mathrm{f}} + \bar{x}'_{\mathrm{m}}(\hat{\beta}_{\mathrm{m}} - \hat{\beta}_{\mathrm{f}}) \\ 0.22 &= -0.01 + 0.23,\end{aligned} \tag{8}$$

where $\overline{\ln(w)}$ is the mean of log wages and the subscripts m and f denote male and female.

The results from estimating (7) and (8) using the regression estimates from Table 2 are listed below each of the equations. The observed difference in the mean of the log wages is equal to 0.22. In terms of wage levels, the trimmed, selection-corrected data indicate that the average male wage is 24% greater than the average female wage. The first component on the right-hand side of (7) and (8) indicates how much of this observed male–female wage gap is explained by differences in observed characteristics such as education and experience. Equation (7) shows that if the wage differential were only due to differences in observable characteristics, the male–female gap in terms of the mean of log wages would be 0.03. Equation (8) indicates that if the differential were only due to differences in characteristics, the gap would be −0.01 indicating that women would actually receive higher mean wages.

The second component of (7) and (8) provides a measure of discrimination.[24] This component shows how much of the observed wage gap is due to the different ways in which men and women are rewarded for their characteristics. Equation (7) indicates that the difference in the mean of log wages due to the different payment structures for men and women is equal to 0.19, while (8) estimates this measure to be 0.23. The estimate from (7) implies that discriminatory labor markets result in men being paid 21% more than women, while (8) indicates that men are paid 25% more.[25]

The difference between (7) and (8) depends on whether the male or female wage structure is used as the basis for comparison. Neumark (1988) shows that the choice is linked to the researcher's expectation of the prevailing wage structure in the

[24] In this model, discrimination includes occupational discrimination.

[25] The estimated percentage change in levels is found in two steps. First, the trimmed, selection-corrected data indicate that the mean male wage is 24% greater than the mean female wage. The second step is to determine what proportion of this 24% difference is attributable to discrimination. The parameter estimates from Eq. (7) indicate that 86% (0.19/0.22) of the difference, or 21%, is attributable to discrimination. The estimates from Eq. (8) indicate that 105% (0.23/0.22) of the difference, or 25%, is attributable to discrimination.

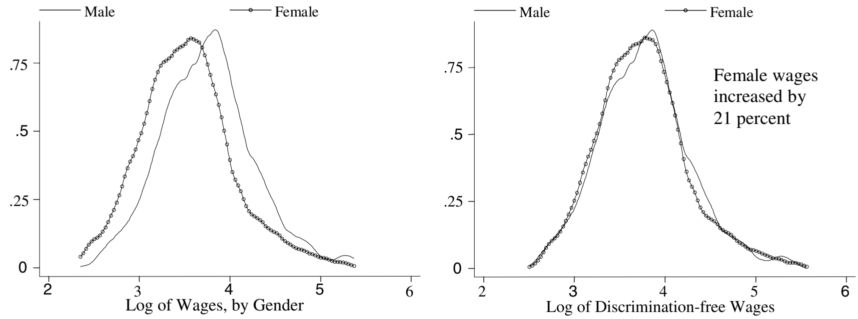

FIG. 2. Distribution of wages by sex, kernel estimate of the density function. Tails of the distribution are trimmed at the 1st and 99th percentile resulting in a sample size of 861 female and 962 male wage earners. The Epanechnikov kernal is used with a bandwidth of 0.09. This width is 90% of the optimal width for estimating the density of a normally distributed variable with the Gaussian kernal.

absence of discrimination.[26] This index number problem affects the interpretation of the estimation results. In Oaxaca's original work, using the male wage structure as the base implied that 53% of the log wage gap was due to discrimination, while using the female wage structure implied that 64% of this gap was due to discrimination.

Neumark suggests that if the researcher believes that women are discriminated against and that the male wage structure would prevail in the absence of discrimination, Eq. (7) provides the appropriate estimates. If the researcher believes that men are somehow receiving a premium resulting from preferential treatment and that the female wage structure would prevail in the absence of discrimination, then Eq. (8) provides the appropriate estimates. In the tight labor market in Bulgaria with falling wages and rising unemployment, the conceptual framework of women being discriminated against, rather than men receiving a wage premium, is somewhat more compelling. For this reason and for the sake of being somewhat conservative, our preferred estimate is that men receive wages that are 21% higher than women resulting from discriminatory labor markets.

The Oaxaca decomposition sheds light on average wage differences due to differences in average characteristics and wage structures. Averages can be misleading if the wage distributions for men and women are dramatically different. The left-hand panel of Fig. 2 presents a kernal density estimate of the log of wages

[26] Neumark (1988) also shows that if employers' preferences for the male–female composition of their labor force are homogeneous of degree zero, the appropriate index for the nondiscriminatory wage structure is the parameter estimates from the pooled sample. Appleton *et al.* (1999) provide a critique of the Neumark decomposition arguing that there is no empirical evidence supporting the assumption of hiring preferences that are homogeneous of degree zero. In the case of Bulgaria, there is some evidence suggesting the contrary. Female participation rates fell from 93 to 81% between 1989 and 1991 (Sziráczki and Windell, 1992). Of course, participation reflects preferences of workers and firms, but this is a dramatic change in the male–female composition of the labor force nonetheless.

for men and women. The two distributions appear to be fairly similarly shaped, except that the female wage distribution is shifted to the left. The right-hand panel presents the same diagram except that female wages have been increased by 21% to compensate for the discriminatory practices of the labor market. This panel makes quite clear that gender discrimination accounts for most of the difference between male and female wages across all points of the distribution.

As a final simulation, the extent to which inequality in wages would decline if women were paid according to the male wage structure is examined. First, an estimate of predicted log wage, called actual predicted wage, is constructed using the separate models for men and women in Table 2. Then log wages are predicted for both men and women using only the parameter estimates for the male model. This procedure provides estimates of what men and women would earn if they were both paid according to the male wage structure and is referred to as the discrimination-free predicted wage.

Any estimate of inequality will be low for both the actual predicted wage and the discrimination-free predicted wage since the log transformation reduces greatly the dispersion, and similarly predicted values will further reduce dispersion. Nonetheless, examining how measures of inequality change when comparing the two measures does provide some insight into the extent to which discrimination worsens wage inequality. The BIHS data indicate that the Gini coefficient and the coefficient of variation for the discrimination-free predicted wage is 18% lower than for the actual predicted wage. Further, the Theil measure of entropy, which weights the tails of the distribution more heavily than the Gini, declines by 34% when actual predicted wages are transformed into discrimination-free predictions.

In summary, the estimation results from the labor supply models indicate that nonwage income variables are important determinants of male labor supply while nonwage income and the number of children are important determinants of female labor supply. Both results support the claim that the decision to participate in the wage sector is identified and that the correction for sample-selection bias is credible. The labor supply estimates are used to trim the sample for estimating separate male and female wage regressions and the wage results demonstrate that there are significant differences in the payment structure for men and women. An Oaxaca decomposition of these differences indicates that the large majority of the difference in wages received by men and women cannot be explained by any of the variables in the model suggesting that sex discrimination may be an important factor in explaining the gender wage gap.

5. CONCLUSION

This paper examines the extent to which discrimination explains the gender wage differential in Bulgaria and the extent to which discrimination affects wage inequality. The results show that elimination of the unexplained portion of the gender wage gap would reduce significantly wage inequality. The distribution of wages would be essentially the same for women and men if all women received an increase in wages of 21% to compensate for the discriminatory practices. Similarly, if women were paid according to the male wage schedule, the level of wage inequality would decline significantly. The Gini coefficient of the log of predicted wages declines by 18% when female wages are predicted using the male wage structure.

The BIHS data indicate that male wages are about 24% greater than female wages. An Oaxaca decomposition of this differential shows that differences in economic characteristics, such as education, experience, and sector of employment, explain very little of this wage differential. The large majority, between 86 and 105%, of this wage differential is explained by what is typically referred to as discriminatory practices. More accurately this difference is explained by the difference in the modeled wage structures for men and women, that is, how men and women are rewarded differently for the same economic characteristics.

While the wage gap in Bulgaria is quite similar to other CEE countries and it is also reasonably similar to the U.S. and other West European countries, the proportion of the gap that can be explained by differences in human capital characteristics and sector of employment is significantly lower. Research from the U.S. suggests that between 30 and 60% of the wage gap is explained by differences in these factors, while this paper shows that between −5 and 14% of the gap can be explained for Bulgaria. This result is also lower than many other CEE countries such as Romania, Hungary, and the Czech Republic, but similar to the results for Russia and Poland. These results suggest that Bulgaria has been less successful than some of the other CEE countries in promoting equal opportunities for women in the labor market. An implication of this research is that promoting social equity will remain a significant barrier for Bulgaria in gaining accession to the European Union, and much effort will be required to promote gender equality in the labor market.

In particular, this research shows that education and experience outcomes do not explain the gender wage gap. Therefore, these should not be the focus of labor market reforms aimed at promoting equality. Similarly, conditioning on economic sector does not explain a significant portion of the gender wage gap; hence policymakers need not focus on comparable worth policies across sectors. Within a given economic sector, women may sort or be sorted into occupation types that are lower paying so that policy must address this issue. Violations of equal pay for equal work within economic sectors may also occur so that policymakers should focus on more vigorously enforcing equal-pay legislation.

REFERENCES

Appleton, Simon, Hoddinott, John, and Krishnan, Pramila, "The Gender Wage Gap in Three African Countries." *Econ. Develop. Cult. Change* **47,** 2:289–312, Jan. 1999.

Becker, Gary, "A Theory of the Allocation of Time." *Econ. J.* **75,** 299:493–517, Sept. 1965.

Beleva, Iskra, Jackman, Richard, and Nenova-Amar, Mariela, "Bulgaria." In Simon Commander and Fabrizio Coricelli, Eds., *Unemployment, Restructuring, and the Labor Market in Eastern Europe and Russia,* pp. 193–232. Washington, DC: The World Bank, 1995.

Blau, Francine, and Kahn, Lawrence, "Gender Differences in Pay." *J. Econ. Perspectives* **14,** 4:75–99, Fall 2000.

Brainerd, Elizabeth, "Women in Transition: Changes in Gender Wage Differentials in Eastern Europe and the Former Soviet Union." *Ind. Lab. Relat. Rev.* **54,** 1:138–162, Oct. 2000.

Commission of the European Communities, *2001 Regular Report on Bulgaria's Progress Towards Accession.* Brussels: Commission of the European Communities, 2001.

Deaton, Angus, *The Analysis of Household Survey*. Baltimore: Johns Hopkins Press, 1997.

Efron, Bradley, and Tibshirani, Robert, *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, Vol. 57, New York: Chapman & Hall, 1993.

Filer, Randall, Jurajda, Štepán, and Ján Plánovský, "Returns to the Market: Valuing Human Capital in the Post Transition Czech and Slovak Republics." *Labour Econ.* **6,** 4:581–593, Nov. 1999.

Heckman, James, "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* **42,** 4:679–694, July 1974.

Honoré, Bo, Kyriazidou, Ekaterini, and Udry, Christopher, "Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons." *J. Econometrics* **76,** 1/2:107–128, Jan./Feb. 1997.

Hurd, Michael, "Estimation in Truncated Samples When There Is Heteroscedasticity." *J. Econometrics* **11,** 2/3:247–258, Oct./Dec. 1979.

Jones, Derek, "The Bulgarian Labour Market in Transition." *Int. Lab. Rev.* **130,** 2:211–226, 1991.

Jurajda, Št pán, "Gender Wage Gap and Segregation in Late Transition." Working Paper No. 182. Prague, Czech Republic: Center for Economic Research and Graduate Education, July 2001.

Laporte, Bruno, and Ringold, Dena, "Trends in Education Access and Financing During the Transition in Central and Eastern Europe." Technical Paper No. 361. Washington, DC: The World Bank, Social Challenges of Transition Series, Apr. 1997.

Linz, Susan, "Gender Differences in the Russian Labor Market. *J. Econ. Issues* **30,** 1:161–185, Mar. 1996.

Milanovic, Branko, *Income, Inequality and Poverty During the Transition from Planned to Market Economy.* Washington, DC: The World Bank, 1998.

Milanovic, Branko, "Explaining the Increase in Inequality During the Transition." *Econ. Transition* **7,** 2:299–341, 1999.

National Statistical Institute, *Statistical Yearbook 1998*. Sofia, Republic of Bulgaria: National Statistical Institute, 1998.

Nelson, Forrest, "A Test for Misspecification in the Censored Normal Model." *Econometrica* **49,** 5:1317–1329, Sept. 1981.

Neumark, David, "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *J. Human Res.* **23,** 3:279–295, Summer 1988.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets." *Int. Econ. Rev.* **14,** 3:693–709, Oct. 1973.

Pagan, Adrian, and Vella, Frank, "Diagnostic Tests for Models Based on Individual Data: A Survey." *J. Appl. Econ.* **4,** 0 (Supplement):s29–s59, Dec. 1989.

Pailhé, Ariane, "Gender Discrimination in Central Europe During the Systemic Transition." *Econ. Transition* **8,** 2:505–535, 2000.

Paternostro, Stefano, and Sahn, David, "Wage Determination and Gender Discrimination in a Transition Economy: The Case of Romania." Policy Research Working Paper No. 2113. Washington, DC: The World Bank, May 1999.

Powell, James, "Least Absolute Deviations Estimation for the Censored Regression Model." *J. Econometrics* **25,** 3:303–325, July 1984.

Powell, James, “Symmetrically Trimmed Least Squares Estimation for Tobit Models.” *Econometrica* **54,** 6:1435–1460, Nov. 1986.

Psacharopoulos, George, “Returns to Investment in Education: A Global Update.” *World Devel.* **22,** 9:1325–1343, Sept. 1994.

Reilly, Barry, “The Gender Pay Gap in Russia during the Transition, 1992–96.” *Econ. Transition* **7,** 1:245–264, 1999.

Sicilian, Paul, and Grossberg, Adam, “Investment in Human Capital and Gender Wage Differences: Evidence from the NLSY.” *Appl. Econ.* **33,** 4:463–471, Mar. 2001.

Singh, Inderjit, Squire, Lynn, and Strauss, John, Eds., *Agricultural Household Models: Extensions, Applications, and Policy*. Baltimore: Johns Hopkins Press, 1986.

Svejnar, Jan, “Labor Markets in the Transitional Central and East European Economies.” In Orley Ashenfelter and David Card, Eds., *Handbook of Labor Economics,* Vol. 3, pp. 2809–2857. Amsterdam/New York: Elsevier, North-Holland, 1999.

Sziráczki, György, and Windell, James, “Impact of Employment Restructuring on Disadvantaged Groups in Hungary and Bulgaria.” *Int. Lab. Rev.* **131,** 4/5:471–496, 1992.

U.S. Census Bureau, International Data Base, Table 001: Total Midyear Population, 2001. Available at http://www.census.gov/ipc/www/idbnew.html. Washington, DC: U.S. Census Bureau, 2001.

World Bank, “Note on Consumption and Income Aggregates.” Mimeo, Bulgaria Integrated Household Survey, Basic Documentation. Washington, DC: The World Bank, 1996.

World Bank, “Bulgarian-Education Modernization Project.” Program Information Document, Report No. PID8735. Washington, DC: The World Bank, Feb. 2000.