

# Appendix to “Skill-biased Technological Change, Earnings of Unskilled Workers, and Crime”

Naci Mocan and Bulent Unel

## 1 The CPS Data

We use the March Current Population Survey (CPS) files from 1978 to 2010 (covering earnings from 1977 to 2009) for full-time workers (those who work 35 or more hours a week) ages 16 to 64. Self-employed people are dropped from the sample, as are allocated earnings observations (using individual earnings allocation flags). In constructing the key variables, we closely follow the previous labor literature on wage inequality (Katz and Murphy, 1992; Krusell et al., 2000; Card and DiNardo, 2002; and in particular, Autor et al., 2008).

### 1.1 Construction of Efficiency-adjusted Labor Inputs

Each individual’s average weekly earning is formed by dividing his annual income (from wages and salaries) by the number of weeks that he worked during the previous year. Earnings are deflated using the state-specific level price deflators from Berry et al. (2000).<sup>1</sup> We make two adjustments for topcoded earnings. First, following Autor et al. (2008) income of workers with top coded earnings is imputed by multiplying the annual topcode amount by 1.5. Second, starting in 1996, topcoded earnings values are assigned the mean of all topcoded earners. In these cases, we simply reassign the topcoded values to all such observations and again multiply by 1.5. Workers whose weekly earnings below \$70 in 2005 dollars are dropped, as are those non-full-year workers (i.e., those who work less than 40 weeks) whose weekly earnings exceed  $1/40^{th}$  the top-coded value of weekly earnings.

We construct the series for high-skill and low-skill labor input and wages as follows. The data in each year in each state are divided into 24 distinct groups characterized by 2 sexes, 4 education categories ( $E \leq 11$ ,  $E = 12$ ,  $13 \leq E \leq 15$ , and  $E \geq 16$ ),<sup>2</sup> and three potential experience categories (0-9, 10-19, 20+ years).<sup>3</sup> Potential experience are calculated as  $\text{Min}\{\text{age} -$

---

<sup>1</sup>Berry et al. (2000) have recently extended their original data set to 2007; and we used the updated series. As an alternative deflator, we also used the US PCE index (2005=1) from the BLS as in Autor et al. (2008). However, results remain qualitatively similar to those reported in the paper.

<sup>2</sup>Commencing in 1992, the Bureau of the Census changed the emphasis of its educational attainment question from years of education to degree receipt. To obtain a comparable educational-attainment data across years, the classification proposed by Jaeger (1997) is followed.

<sup>3</sup>This taxonomy is the same as in Autor et al. (2008) and many others. However, due to limitations in the availability of state-level data, we consider a higher level of divisions. Since there are 50 states, the above taxonomy divides the annual data into 1200 groups.

years of schooling–6, age–16} following Autor et al. (2008). In constructing labor supply data, for each state/region, the total weeks for each group is normalized by the sum of total weeks worked over all groups so that weeks for each group in each year are expressed as a fraction of total annual weeks. In calculating each group’s average weekly earnings, earnings are weighted by the product of the corresponding CPS sampling weight and weeks worked.

We assume that the high-skill labor class consists of college or college-plus workers and the workers with some college; and the low-skill labor class consists of those who have no college education. Groups within a class are assumed to be perfect substitutes and we use group relative weekly earnings of full-time workers as weights for the aggregation of labor inputs into skilled and unskilled classes. Standard in this literature is the assumption that relative wages equal relative efficiencies of labor. More specifically, following Autor et al. (2008), we choose the group that contains male workers with less than 12 years of education and with less than 10 years of potential experience as the base group. A relative wage measure is then constructed by dividing each group’s average weekly earnings by the average weekly earning of the base group. The relative efficiency index measure for each group,  $q_g$ , is computed as the arithmetic mean of the relative wage measures in that group over 1977 to 2009. Then the total efficiency-adjusted labor input in each class is given by

$$H_t = \sum_{g \in G_H} n_{gt} q_g, \quad L_t = \sum_{g \in G_L} n_{gt} q_g,$$

where  $n_{gt}$  represents the total labor weeks used in production by group  $g$  in year  $t$ .<sup>4</sup>

Since  $H$  and  $L$  are efficiency-adjusted labor inputs, the corresponding earnings  $W_H$  and  $W_L$  in equation (5) are also efficiency-adjusted. Following Krusell et al. (2000), they are calculated as

$$W_{Ht} = \sum_{g \in G_H} n_{gt} w_{gt} / H_t, \quad W_{Lt} = \sum_{g \in G_L} n_{gt} w_{gt} / L_t,$$

where  $w_{gt}$  represents the average weekly earnings of group  $g$  in year  $t$ .

As an alternative measure of earnings, we adjust for the *composition* of labor input so that the average weekly earnings of high-skill and low-skill workers are not mechanically affected by shifts in the experience, gender composition, or average level of completed schooling (Autor et al., 2008). The composition index for each group,  $n_g$ , is computed as the arithmetic mean of  $n_{gt}$  over 1977 to 2009. Then the composition-adjusted weekly wages are given by

$$W_{Ht} = \sum_{g \in G_H} n_g w_{gt} / \sum_{g \in G_H} n_g, \quad W_{Lt} = \sum_{g \in G_L} n_g w_{gt} / \sum_{g \in G_L} n_g.$$

---

<sup>4</sup>As indicated above, the total labor weeks for each group is expressed as a fraction of total annual weeks.

## 1.2 Industry-State Level Analysis

Construction of the key variables at industrial level follows the same steps. In each state, we divide industries in three groups: manufacturing, service, and other (agriculture + mining + construction). Annual data in each industry in each state are divided into 8 distinct groups characterized by 2 sexes and 4 education categories ( $E \leq 11$ ,  $E = 12$ ,  $13 \leq E \leq 15$ , and  $E \geq 16$ ). In calculating each group's average weekly earnings, earnings are weighted by the product of the corresponding CPS sampling weight and weeks worked. High-skill and less-skill labor classes are the same as above; and while aggregating labor inputs into skilled and unskilled classes, we choose the group that contains male workers with less than 12 years of education and with less than 10 years of potential experience as the base group.

## 1.3 Selection into Labor Force

Table A-1 presents the results obtained from estimation of the selection equation and the market wage equation. The two equations are estimated jointly using maximum likelihood. The results are consistent with those obtained from typical wage studies where having at least a high school degree and being male have a positive impact on wages, but being Black has a negative impact on wages in comparison to whites. State income and living in urban areas are positively correlated with wages. Non-labor income has a negative impact on the propensity to work in the legal labor market; the same is true for state income and the unemployment rate. Higher education makes the person more likely to participate in the legal labor market. Blacks and Hispanics have lower propensities to participate. The correlation between the errors of the selection and wage equation is negative as it is sometimes found in other studies (Wright and Ermisch 1991; Steinberg 1989). This is possible in variety of circumstances. For example a negative correlation emerges if the variance of market wages is smaller than the covariance between market wages and reservation wages (Ermisch and Wright 1994).

**Table A1.** NLSY97 Data; Selection into Labor Force, and Market Wages

Variable	Selection Equation	Wage Equation
Non Labor Income	-0.0001** (0.00005)	
Household Size	-0.001 (0.005)	-0.007*** (0.002)
High School +	0.089*** (0.022)	0.071*** (0.007)
Male	0.014 (0.016)	0.047*** (0.006)
Black	-0.272*** (0.022)	-0.022*** (0.008)
Hispanic	-0.131*** (0.024)	-0.009 (0.008)
Age	0.181*** (0.008)	0.023*** (0.003)
Urban	-0.017 (0.020)	0.029*** (0.007)
Married	-0.235*** (0.044)	0.072*** (0.014)
Separated	-0.076 (0.155)	-0.073 (0.075)
Divorced	-0.345** (0.146)	0.025 (0.041)
Children	-0.160*** (0.022)	0.013* (0.007)
State Inc/Capita	-0.026 (0.017)	0.016** (0.007)
Unemployment	-0.040** (0.019)	0.009 (0.007)
% 15-24 Yr. Old	-0.081** (0.035)	0.034** (0.013)
% Black population	-0.059* (0.035)	0.030* (0.015)
Observations	55,037	55,037
$\rho$	-0.811	

*Notes:* Robust standard errors, clustered at the individual level, are in parentheses. \* signifies statistical significance at the 10% level; \*\* at 5% level, and \*\*\* at the 1% level or less. State alcohol consumption, state and year dummies are included in all regressions.

**Table A2.a:** NLSY97 Data, Instrumental Variables Regressions—Regional Instrument

	Larceny/ Car Theft/ Robbery I	Larceny II	Car Theft III	Stolen Property IV	Selling Drugs V	Selling Hard Drugs VI	Burglary VII	Robbery VIII
Log Wage	-0.218*** (0.060)	-0.196*** (0.059)	-0.028 (0.019)	-0.057 (0.037)	-0.055 (0.048)	-0.059** (0.030)	-0.028 (0.027)	0.011 (0.015)
Arrest Rate	-0.002 (0.001)	-0.003* (0.002)	-0.008*** (0.003)	-0.000 (0.000)	-0.001 (0.002)	0.002 (0.001)	-0.007*** (0.002)	-0.003 (0.003)
High School +	0.002 (0.008)	-0.001 (0.008)	0.003 (0.003)	-0.002 (0.005)	-0.013** (0.006)	-0.003 (0.004)	0.003 (0.004)	-0.001 (0.002)
Hshld Inc	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Age	-0.015*** (0.004)	-0.012*** (0.004)	-0.001 (0.001)	-0.003 (0.002)	-0.003 (0.003)	-0.001 (0.002)	-0.003* (0.002)	-0.001* (0.001)
Gun	0.098*** (0.010)	0.086*** (0.010)	0.042*** (0.006)	0.087*** (0.008)	0.094*** (0.009)	0.063*** (0.007)	0.053*** (0.007)	0.036*** (0.004)
Heavy Drinking	0.003*** (0.001)	0.002*** (0.001)	0.001*** (0.000)	0.002*** (0.001)	0.005*** (0.001)	0.003*** (0.001)	0.002*** (0.000)	0.001** (0.000)
Marijuana Use	0.003*** (0.000)	0.003*** (0.001)	0.000 (0.000)	0.002*** (0.000)	0.011*** (0.000)	0.003*** (0.000)	0.001** (0.000)	0.000** (0.000)
State Inc/Cap	-0.003 (0.004)	-0.002 (0.004)	-0.001 (0.001)	-0.001 (0.002)	-0.000 (0.002)	-0.001 (0.001)	-0.002 (0.001)	0.000 (0.001)
% Black	0.011 (0.009)	0.016* (0.009)	-0.003 (0.003)	0.001 (0.004)	-0.002 (0.006)	0.002 (0.004)	-0.000 (0.003)	0.001 (0.001)
Observations	54,411	47,479	47,479	54,398	54,386	54,382	47,479	54,658

*Notes:* Robust standard errors, clustered at the individual level, are in parentheses. \* signifies statistical significance at the 10% level; \*\* at 5% level, and \*\*\* at the 1% level or less. Urban residence, marital status, number of children, share of population aged 15-24, state and year dummies and individual fixed-effects are included in all regressions. First-stage F-statistics for the excluded instrument in various crime regressions are greater than 110.

**Table A2.b:** NLSY97 Data, Instrumental Variables Regressions, Wage Increase and Decrease –Regional Instrument

	I	II	III	IV	V	VI	VII	VIII
	Larceny/ Car Theft/ Robbery	Larceny	Car Theft	Stolen Property	Selling Drugs	Selling Hard Drugs	Burglary	Robbery
Log Wage <sup>+</sup>	-0.306** (0.131)	-0.187* (0.112)	-0.041 (0.037)	-0.091 (0.079)	0.013 (0.101)	-0.114* (0.064)	-0.014 (0.046)	0.041 (0.032)
Log Wage <sup>-</sup>	-0.322** (0.140)	-0.195* (0.119)	-0.041 (0.039)	-0.097 (0.084)	0.014 (0.107)	-0.120* (0.069)	-0.013 (0.049)	0.044 (0.034)
Arrest Rate	-0.003 (0.002)	-0.003 (0.003)	-0.003 (0.004)	-0.000 (0.000)	-0.000 (0.002)	0.000 (0.002)	-0.001 (0.003)	0.002 (0.004)
High School +	-0.006 (0.007)	-0.008 (0.007)	0.003 (0.002)	-0.002 (0.004)	-0.012** (0.005)	-0.003 (0.004)	0.001 (0.003)	-0.000 (0.001)
Hshld Inc	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Age	-0.019*** (0.005)	-0.018*** (0.005)	-0.001 (0.001)	-0.006** (0.003)	-0.007* (0.004)	-0.000 (0.003)	-0.004** (0.002)	-0.002 (0.001)
Gun	0.095*** (0.012)	0.077*** (0.012)	0.032*** (0.006)	0.081*** (0.009)	0.097*** (0.010)	0.060*** (0.008)	0.043*** (0.007)	0.035*** (0.005)
Heavy Drinking	0.003*** (0.001)	0.002*** (0.001)	0.001** (0.000)	0.002*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.001*** (0.000)	0.000 (0.000)
Marijuana Use	0.003*** (0.001)	0.003*** (0.001)	0.000 (0.000)	0.001*** (0.000)	0.010*** (0.001)	0.002*** (0.000)	0.001* (0.000)	0.000* (0.000)
State Inc/Cap	-0.007* (0.004)	-0.007 (0.004)	-0.001 (0.001)	-0.001 (0.002)	0.001 (0.003)	0.002 (0.002)	-0.001 (0.002)	-0.000 (0.001)
% Black	0.010 (0.011)	0.015 (0.011)	-0.002 (0.003)	-0.001 (0.005)	-0.008 (0.008)	0.002 (0.005)	0.001 (0.004)	-0.000 (0.002)
N	43,043	36,565	36,566	43,041	43,028	43,025	36,566	43,234

*Notes:* Robust standard errors, clustered at the individual level, are in parentheses. \* signifies statistical significance at the 10% level; \*\* at 5% level, and \*\*\* at the 1% level or less. Urban residence, marital status, number of children, share of population aged 15-24, state and year dummies and individual fixed-effects are included in all regressions. First-stage F-statistics for the excluded instrument in various crime regressions are greater than 260.