Online Appendix to Card, Heining, and Kline (2013)

January 22, 2013

Contents

In this appendix, we provide additional details on the construction of our sample, the nature of our imputation procedure, and provide some supplementary results.

1 Sample Construction and Processing

1.1 Creation of Wage Data

Our basic data source is the Integrated Employment Biography (IEB) database. The IEB consists of information on employment spells at a given establishment (or firm) within a calendar year, taken from notices of employment filed by the employer. Each notice of employment has a beginning date (e.g., January 1), an end date (e.g., December 31), the average daily wage earned by the employee (censored at the Social Security maximum earnings level), indicators for the legal status of the job (including whether the job is full time or part time and whether the job is a "marginal job" subject to reduced Social Security taxes), as well as information on the gender, birth date, educational qualification and occupation of the worker, and the industry and geographic location of the establishment.

We process the data in two steps. First, we collapse all spells that are recorded as full-time jobs at the same employer in a given year into a single person-firm-year record, summing total earnings at each employer. Row 1 of panels A and B in Appendix Table 1 shows the numbers of spell records in the IAB data file for full-time male and female employees age 20-60 working in non-marginal jobs in 1985 (the first year of our sample), 1997 (the middle year of our sample) and 2009 (the last year). Row 2 in each panel shows the number of person-firm-year records: on average there are about 1.06 spells per person-firm-year in the early years of our sample, rising slightly to 1.11 spells per person-firm-year in 2009 (see row 4).

In the second step we select one observation per person per year by selecting the person-firm record with the highest total earnings in a given year (and excluding any observation with a daily wage $<$ 10 Euros). On average each person has about 1.1 different employers per year, with only a small upward trend over our sample period (see row 5 in each panel). Line 3 in panels A and B shows the numbers of person-year observations for full time men and women in our final data set.

1.2 Education

Education levels in the IEB are coded into 6 categories, plus a missing or undetermined category. We group these into 5 classes: (1) missing; (2) primary/lower secondary or intermediate school leaving certificate, or equivalent, with no vocational qualification; (3) primary/lower secondary or intermediate school leaving certificate, or equivalent, with a vocational qualification; (4) upper secondary school certificate ("Arbitur") with or without a vocational certificate; (5) degree from Fachhochschule or university. For simplicity we refer to the third category as "apprentices" and the fifth category as "university graduates". For an individual who is observed in multiple notifications from the same employer in the same year we assign the highest education category for that person-firm-year observation. Within each job spell we assign the modal education category observed for an individual in the years he is at the same job.

1.3 Occupation and Industry

In the IEB data, each job notification includes information on occupation and industry. For an individual who is observed in multiple notifications from the same employer in the same year we assign the highest occupation category and the highest industry category for that person-firm-year observation. Within job spells the industry code is constant in 97% of spells. In the remaining cases we assign the highest industry category observed over the years of the job. We do not assign a fixed occupation code to job spells.

1.4 Tobit imputations

As illustrated in Table 1, roughly 10 percent of person-year observations for male workers and 1-2 percent of the observations for female workers are censored at the Social Security maximum. We follow Dustmann et al. (2009) and fit a series of Tobit models to log daily wages. We then impute an uncensored value for each censored observation using the estimated parameters of these models and a random draw from the associated (left- censored) distribution.[1]

Since we are fitting models that include both a person and year effect, we want the imputation model to reflect individual and job-specific components of the wage. We therefore construct, for each individual in each year, the mean of his log wage in all other periods, and the fraction of other years that the individual's wage is censored. For individuals who are only observed in one year, we set the mean log wage in other years to the sample mean, and the fraction of censored wages in other years equal to the sample mean, and include a dummy in the model for those who are observed only once. We also construct the mean log wage for the individual's co-workers in the current year (i.e., the "leave out mean" of log wages at his employer) and the fraction of co-workers who are censored in the current year (the "leave out mean" of the censoring rate at his employer). For individuals who work at an establishment with only 1 full time male employee we set the mean log wage for co-workers equal to the sample mean, and the fraction of co-workers with censored wages equal to the sample mean, and include a dummy in the model for employees of 1-worker firms.

We then fit a series of 500 Tobit models separately by year, education (5 values: missing; no qualification; apprenticeship; some post secondary; and university graduate), and 10 year age range (20-29; 30-39; 40-49; 50-60), including the following variables: age, mean log wage in other years, fraction of censored wages in other years, number of full time male employees at the current firm and its square, dummy for 11 or more employees, mean years of schooling and fraction of university graduates at the current firm, mean log wage of co-workers and fraction of co-workers with censored wages, dummy for individuals observed only 1 year between 1985 and 2009, dummy for employees of 1-worker firm. Appendix Table 2 shows

[1] Specifically, we impute an upper tail as follows. Suppose that the estimated Tobit model for y (the log of wages) has $y \sim N(X'\beta, \sigma)$, and consider a censored observation, such that $y \geq c$, where c is the censoring point. Let $k = \Phi[(c - X'\beta)/\sigma]$, where Φ represents the standard normal density, and let $u \sim U[0,1]$ represent a uniform random variable. Then we impute an uncensored value for y as: $y^u = X'\beta + \sigma\Phi^{-1}[k + u \times (1-k)]$.

the coefficient estimates for models for 40-49 year old apprentices in 1985, 1997, and 2009.

1.5 Validation Exercise

To evaluate the quality of the approximation to the upper tail provided by our Tobit specification, we performed a validation exercise in which we artificially censor the upper tail of wages for a group of workers with very low censoring rates, then fit Tobit models (with the same explanatory variables as in our main procedure) and stochastically impute the upper tail of wages. We then compare the standard deviation of wages for the original sample with the standard deviation from the censored/imputed sample. We use male workers age 20-29 with an apprenticeship education in years from 1990 to 2009 as the population of interest. These workers have an average censoring rate over the 20 year period of 0.7%. We select artificial censoring points so that 10, 20, 30, or 40 percent of workers are censored in each year, and fit separate Tobit models by year for each censored subsample.

Appendix Figure 1 shows actual standard deviation of wages for the test sample, which rises from 0.296 to 0.348 (an increase of 0.052) between 1990 and 2009, as well as the standard deviations from the censored/imputed samples with differing censoring rates. The standard deviation of the imputed series is uniformly higher than the standard deviation of the raw data, with a larger upward bias at higher censoring rates. For example, when the censoring rate is 40%, the estimated standard deviation is upward biased by about 0.04 (or 13%) in every year. Fortunately, the upward bias is relatively constant, so the trend in the dispersion of wages is very similar whether we use the raw data or any of the censored/imputed series. This leads us to conclude that our Tobit imputation procedure performs relatively well, even for subgroups with very high censoring rates.

A concern about our imputation procedure is that it may alter the relative share within versus between establishment variation. To check this, we fit linear regression models with year dummies and establishment effects to observations from 2002 to 2009 (the same time span as our fourth sample interval), using the raw wage data for 20-29 year old men with an apprenticeship education, and the censored/imputed data. The sample has 8,426,930 person-year wage observations on employees at 668,285 establishments. The R-squared coefficients from the different samples were as follows:

raw data:	0.721
10% censored	0.719
20% censored	0.718
30% censored	0.714
40% censored	0.707

We conclude that the imputation procedure successfully maintains the relative share of the variance of wages attributable to within-establishment variation, even at very high censoring rates.

2 Computational Methods

Because our dataset is very large and identification of the establishment effects derives entirely from movers, we conducted estimation in two steps. First, in each interval, we extracted the sample of workers who switched establishments over the relevant time period. We fit the model to this sample of movers and recovered the estimated vector of establishment effects $\widehat{\psi}$ along with the coefficients $\widehat{\beta}$ corresponding to the time varying covariates x_{it}.[2] Then, for each worker who stayed at the same establishment over the sample interval, we computed an estimate of his person effect as follows:

$$\widehat{\alpha}_i = \frac{1}{T_i} \sum_t \left(y_{it} - \widehat{\psi}_{\mathbf{J}(i,t)} - x'_{it}\widehat{\beta} \right)$$

where T_i is the number of periods that individual i is observed in the sample interval. Our root mean squared error calculations were conducted by reducing the degrees of freedom by one for each connected stayer mean estimated.[3]

Our estimation tasks were performed in Matlab. Code for our analysis is available online. We used a variant of the depth first search algorithm implemented in the open source matlabBGL package to find the largest connected set of establishments in each data interval. The design matrices $Z \equiv [D, F, X]$ were stored as sparse matrices. To compute the least squares solutions we solved the normal equations in equation (3) of the paper using Matlab's preconditioned conjugate gradient routine (see Shewchuck (1994) for a lucid introduction). To speed the process we used an incomplete Cholesky factorization of $Z'Z$ as the preconditioner with threshold dropping tolerance of 0.01.

[2] This yields an inefficient estimator of β. However, in a sample of roughly 90 million observations, precision is not a major concern. A separate issue is that our two step procedure only ensures orthogonality between the AKM residuals $\widehat{r}_{it}$ and x_{it} in the sample of movers. In practice, the correlation in the sample of establishment stayers between $x'_{it}\widehat{\beta}$ and the AKM residuals is very small in each interval, with the largest correlation occuring in interval 3 and amounting to approximately $-.01$.

[3] That is we used the formula $RMSE = \sqrt{\frac{SSR}{dof}}$ where SSR is the sum of squared residuals across all person-year observations in the interval and $dof = N^* - N - (J-1) - rank\,(X)$, where N^* is the number of person year observations including the stayers, N is the number of connected individuals including the stayers, and J is the number of connected establishments.

3 Bias in the estimated covariance matrix of person and establishment effects

It is well known that sampling errors in the estimated person and establishment effects may lead to inflated estimates of the standard deviation of each component and negatively biased estimates of the covariance between the person and establishment effects. This has led some authors (e.g., Andrews et al., 2008) to propose parametric bias corrections to the estimated components. To illustrate the logic of such an approach, and the difficulties involved, we denote the population quantities of interest as:

$$\begin{aligned}
\sigma^2_{D\alpha} &\equiv \frac{1}{N^*-1}\alpha' D' Q_1 D\alpha \text{ (Variance of person effects)} \\
\sigma^2_{F\psi} &\equiv \frac{1}{N^*-1}\psi' F' Q_1 F\psi \text{ (Variance of establishment effects)} \\
\sigma_{D\alpha,F\psi} &\equiv \frac{1}{N^*-1}\psi' F' Q_1 D\alpha \text{ (Covariance of person and establishment effects)}
\end{aligned}$$

where $Q_1 \equiv I - 1\left(1'1\right)^{-1} 1'$ is a symmetric demeaning matrix.

OLS estimation of (2) yields a coefficient vector:

$$\widehat{\xi} = \xi + \left(Z'Z\right)^{-1} Z'r$$

with $E\left[\left(Z'Z\right)^{-1} Z'r\right] = 0$. The sampling variance of this vector can be written:

$$\begin{aligned}
V_{\widehat{\xi}} &\equiv E\left[\left(\widehat{\xi}-\xi\right)\left(\widehat{\xi}-\xi\right)'\right] \\
&= \left(Z'Z\right)^{-1} Z'\Omega Z \left(Z'Z\right)^{-1}
\end{aligned}$$

where $\Omega \equiv E\left[rr'\right]$ is the $N^* \times N^*$ variance covariance matrix of the errors.

The sample analogues to the population quantities can be expressed in terms of the following quadratic forms:

$$\begin{aligned}
\widehat{\sigma}^2_{D\alpha} &\equiv \frac{1}{N^*-1}\widehat{\xi}' A_{D\alpha}\widehat{\xi} \\
\widehat{\sigma}^2_{F\psi} &\equiv \frac{1}{N^*-1}\widehat{\xi}' A_{F\psi}\widehat{\xi} \\
\widehat{\sigma}_{D\alpha,F\psi} &\equiv \frac{1}{N^*-1}\widehat{\xi}' A_{D\alpha,F\psi}\widehat{\xi}
\end{aligned}$$

where $A_{D\alpha} \equiv \begin{bmatrix} D'Q_1D & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_{F\psi} \equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & F'Q_1F & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_{D\alpha,F\psi} \equiv \begin{bmatrix} 0 & F'Q_1D & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Unbiasedness of OLS and standard results on quadratic forms imply that for any matrix A,

$$E\left[\widehat{\xi}' A\widehat{\xi}\right] = \xi' A\xi + tr\left(AV_{\widehat{\xi}}\right).$$

Therefore the bias in our estimates of the variance components corresponds to the trace term in the above expression, which in turn depends critically upon $V_{\widehat{\xi}}$. Previous work has focused on evaluating this bias expression under the assumption that the r are independent and identically distributed in which case $\Omega = I\sigma^2$ and $V_{\widehat{\xi}} = (Z'Z)^{-1}\sigma^2$.

Unfortunately, the bulk of the literature on earnings dynamics (MaCurdy, 1982; Abowd and Card, 1989; Meghir and Pistaferri, 2004) suggests a substantially more complicated error structure of earnings with complex forms of temporal dependence and heteroscedasticity. Errors in modeling the structure of Ω will induce errors in estimation of $V_{\widehat{\xi}}$ which is why, at least since the work of White (1980), economists have sought robust variance estimates that don't rely upon estimation of all elements of Ω. Unfortunately, robust variance estimation is not possible in our setting because the estimates $\widehat{\xi}$, while unbiased, are not consistent.

In unreported results we have attempted parametric corrections allowing for a match component and a moving average component to the errors r. These corrections yielded small changes in the estimated variance-covariance matrix $\widehat{V}_{\widehat{\xi}}$ and had trivial effects on the trends of the various components. In sampling experiments we found the corrections to provide a poor guide to the degree of bias created by working with subsamples of the data. We suspect this is because our model for the errors is insufficiently rich – a problem we are unlikely to be able to solve in a convincing way. For this reason, our decompositions in section 7 of between group means are of particular interest because these results are based upon group averages involving tens of thousands (or in some cases millions) of observations, in which case biases due to sampling error become largely irrelevant.

References

1. Abowd, John and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." Econometrica 57(2): 411-445.

2. Andrews, M.J., L. Gill, T. Schank, and R. Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" Journal of the Royal Statistical Society 171(3): 673-679.

3. Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg. 2009. "Revisiting the German Wage Structure." Quarterly Journal of Economics 124(2): 843-881.

4. Macurdy, Thomas. 1982. "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." Journal of Econometrics 18(1): 83-114.

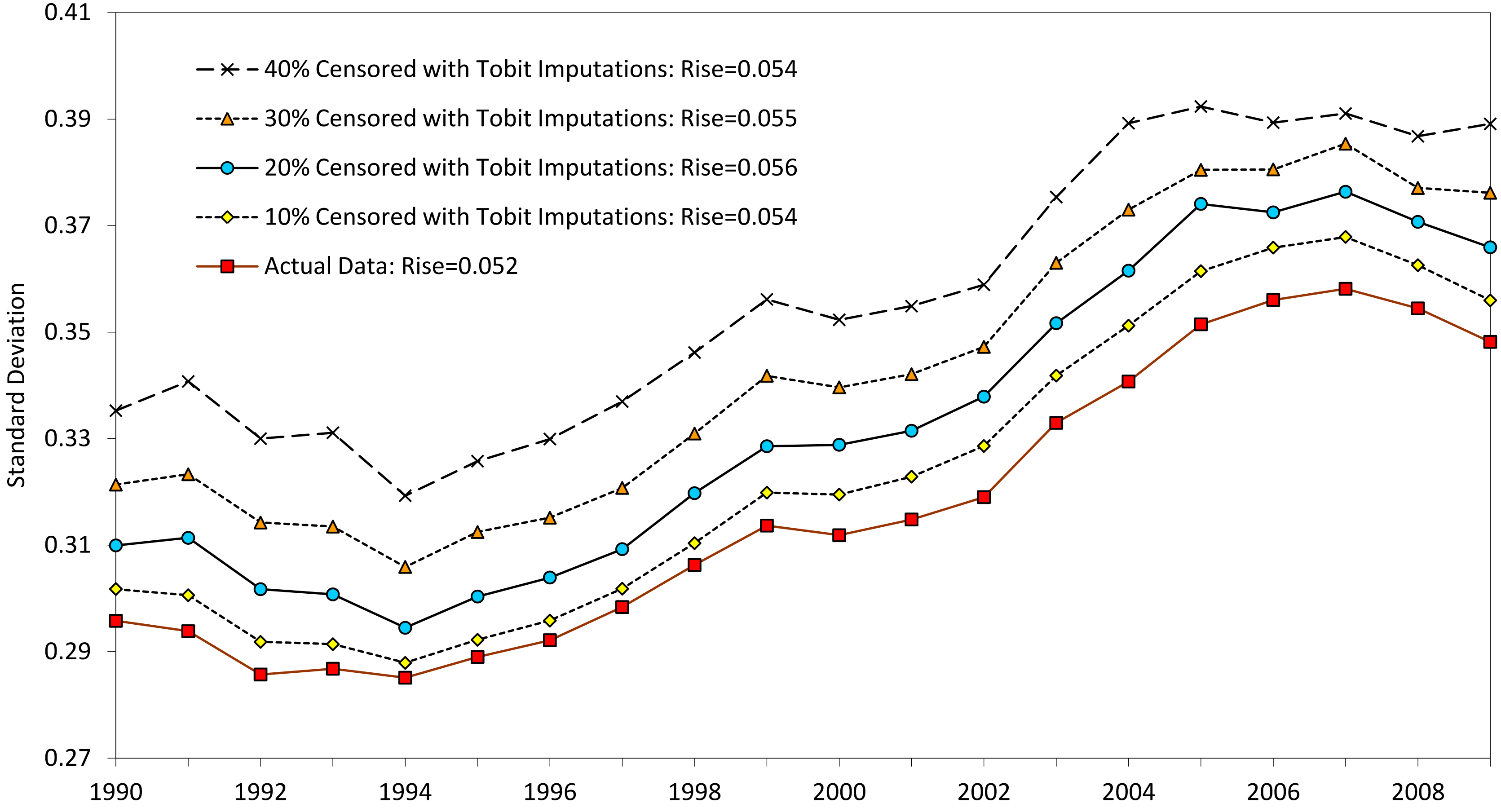
5. Meghir, Costas and Luigi Pistaferri. 2004. "Income Variance Dynamics and Heterogeneity." Econometrica 72(1): 1-32.

6. Shewchuck, Jonathan Richard. 1994. "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain: Edition $1\frac{1}{4}$" Working paper. School of Computer Science, Carnegie Mellon University.

7. White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity." Econometrica 48(4): 817–838.

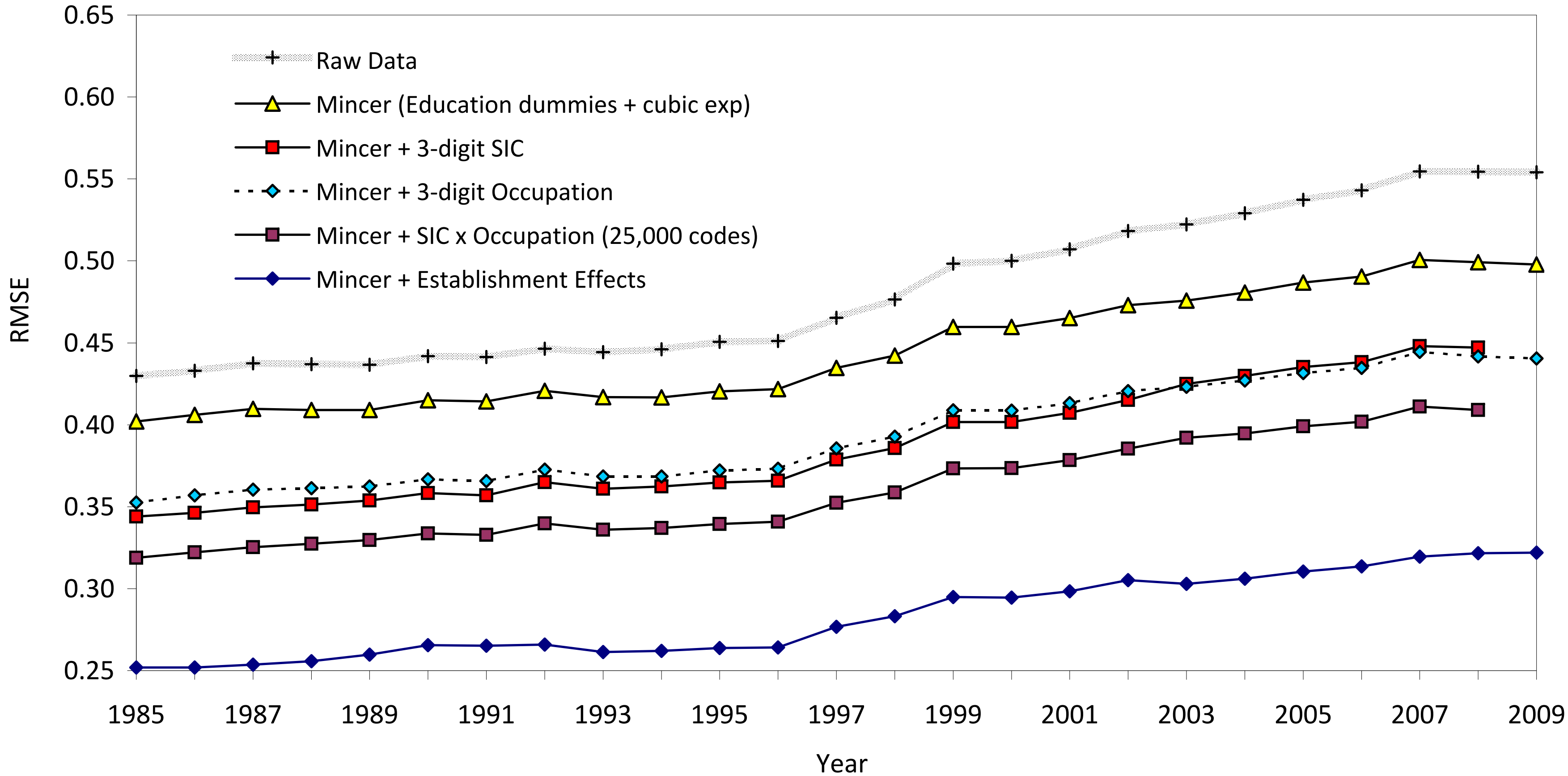
4 Appendix Figures and Tables

Appendix Figure 1: Trends in Standard Deviations of Log Wages -
Male Apprentices Age 20-29, Actual and Artificially Censored/Imputed Data



Note: Actual data has censoring rate of 0.5% or less in every year. Data are artificially censored at real wages levels yielding average censoring rates of 10,20,30, or 40% over the entire sample period. Then Tobit models are fit separately by year, with same covariates as used in main imputation model, and upper tail observations are randomly imputed using same procedure as in main imputation model.

Appendix Figure 2: Raw and Residual Standard Deviations from Alternative Wage Models for Full Time Females



Notes: See notes to Figure 4. Figure shows measures of dispersion in actual and residual real daily wage for full time female workers. Residual wage is residual from linear regression model. "Mincer" refers to model with dummies for education categories and cubic in experience, fit separately in each year. Other models add additional controls as indicated.

Appendix Table 1a: Job Spells and Main Jobs in IEB Data Base

	1985	1997	2009
A. Male Workers			
1. Number of Full Time Job Spells (age 20-60, non-marginal jobs)	13,987,548	14,919,079	14,911,559
2. Number of unique person-firm-year observations	13,181,917	13,968,286	13,456,305
3. Number of person-year observations (highest-paying job only)	11,980,159	12,661,995	12,104,223
4. Average number of spells per job in year	1.06	1.07	1.11
5. Average number of jobs per year	1.10	1.10	1.11
B. Female Workers			
1. Number of Full Time Job Spells (age 20-60, non-marginal jobs)	6,965,926	7,937,037	8,142,682
2. Number of unique person-firm-year observations	6,642,114	7,408,237	7,328,008
3. Number of person-year observations (highest-paying job only)	6,068,863	6,758,622	6,566,429
4. Average number of spells per job in year	1.05	1.07	1.11
5. Average number of jobs per year	1.09	1.10	1.12

Notes: each "job spell" represents a notification of employment in the IEB data base. For each gender, the entry in row 1 is the number of such notifications for full time, non-marginal jobs held by men age 20-60. Row 2 shows the number of unique person-firm-year observations after collapsing multiple spells at the same employer in the same year. Row 3 shows the number of person observations after selecting the person-firm-year observation with the highest total earnings in the year as the "main job" in a given year. Row 4 gives the ratio of row 1 to row 2. Row 5 gives the ratio of row 2 to row 3.

Appendix Table 2: Selected Tobit Models for Male Apprentices, Age 40-49

	1985	1997	2009
Moments of Unadjusted Log Wage Data:			
Mean	4.382	4.4612	4.434
Std. Deviation	0.2677	0.3104	0.3563
Fraction Censored	0.155	0.107	0.0726
Parameter Estimates from Tobit model:			
Intercept	0.539	-0.108	-0.194
	(0.030)	(0.020)	(0.003)
Age/10	0.026	-0.009	-0.103
	(0.002)	(0.002)	(0.003)
Fraction of person's other wage	0.268	0.270	0.699
observations censored	(0.003)	(0.005)	(0.011)
Mean of log wage for person	0.693	0.909	0.819
in other years	(0.003)	(0.003)	(0.004)
Dummy for firm size > 10	-0.006	-0.006	0.005
	(0.002)	(0.002)	(0.002)
Fraction of workers at firm with	-0.045	-0.013	0.007
university degree*	(0.012)	(0.011)	(0.014)
Mean years of schooling of workers	-0.014	-0.010	-0.007
at firm*	(0.002)	(0.002)	(0.002)
Fraction of co-workers with	0.093	0.000	0.135
censored wage**	(0.007)	(0.007)	(0.010)
Mean log wage of co-workers**	0.187	0.163	0.374
	(0.004)	(0.004)	(0.004)
Dummy if person observed in	-0.398	-0.935	-0.808
only 1 year	(0.005)	(0.013)	(0.016)
Dummy if firm has only 1	-0.223	-0.178	-0.386
worker	(0.005)	(0.005)	(0.006)
Estimated scale parameter	0.147	0.135	0.195
	(0.001)	(0.001)	(0.001)
Sample size	62,889	58,392	65,904

Notes: standard errors in parentheses. Table entries are coefficient estimates from Tobit models fit to log real daily wages, with censoring at the Social Security maximum contribution rate. Models also include firm size (number of current-year full time male employees) and its square.

*Characteristics of full-time male (non-marginal) employees at the same firm.

**Statistic is calculated for full-time male employees at the same firm, excluding the individual of interest. For employers with one firm, statistic is set to mean.

Appendix Table 3: Estimation Results for AKM Model, Fit by Interval for Male Apprentices

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Person and Establishment Parameters:				
Number person effects	10,128,342	10,346,742	9,288,956	8,145,059
Number establishment effects	1,008,959	1,078,911	1,093,438	990,608
Summary of Parameter Estimates:				
Std. dev. of person effects (across person-year obs.)	0.241	0.249	0.271	0.285
Std. dev. of establ. effects (across person-year obs.)	0.152	0.159	0.170	0.202
Std. dev. of Xb (across person-year obs.)	0.121	0.080	0.073	0.064
Correlation of person/est effects (across person-year obs.)	-0.035	0.008	0.048	0.098
Correlation of person effects/Xb (across person-year obs.)	0.050	0.075	-0.029	0.020
Correlation of establ. effects/Xb (across person-year obs.)	0.064	0.064	0.048	0.082
RMSE of AKM residual	0.110	0.109	0.110	0.110
Adjusted R-squared	0.887	0.891	0.901	0.919
Comparison Match Model				
RMSE of Match model	0.096	0.095	0.094	0.094
Std. Dev of Match Effect*	0.054	0.052	0.056	0.057
Std. Dev. Log Wages	0.328	0.329	0.349	0.388
Sample size	54,993,845	56,701,812	51,031,280	50,700,611

Notes: see notes to Table 4. Models reported here are estimated for subsample of full time male workers with apprenticeship training only. Xb includes year dummies and quadratic and cubic terms in age (total of 8 parameters in intervals 1-3, 9 in interval 4). Match model includes Xb and separate dummy for each job (person-establishment pair).

*Standard deviation of match effect estimated as square root of difference in mean squared errors between AKM model and match effect model.

Appendix Table 4: Decompositions of Rise in Wage Inequality for Apprentices

	Interval 1		Interval 4		Change from Interval 1 to 4	
	Var. Component (1)	Share of Total (2)	Var. Component (3)	Share of Total (4)	Var. Component (5)	Share of Total (6)
Total variance of log wages	0.108	100.0	0.150	100.0	0.043	100
<u>Components of Variance:</u>						
Variance of person effect	0.058	53.7	0.081	54.2	0.024	55
Variance of establ. effect	0.023	21.3	0.041	27.1	0.018	42
Variance of Xb	0.015	13.6	0.004	2.7	-0.011	-25
Variance of residual	0.010	8.9	0.010	6.6	0.000	1
2cov(person, establ.)	-0.003	-2.4	0.011	7.5	0.014	32
2cov(Xb, person+establ.)	0.005	4.9	0.003	1.9	-0.002	-6
Counterfactuals for Variance of log wages: *						
1. No rise in correl. of person/estab. effect	0.108		0.143		0.035	82
2. No rise in var. of estab. effect	0.108		0.131		0.023	55
3. Both 1 and 2	0.108		0.125		0.018	42

Notes: see notes to Table 5. Calculations based on estimated AKM models summarized in Appendix Table 3.

*Counterfactual 1 computes the counterfactual rise in variance assuming the correlation between the person and establishment effects remains at its interval 1 value -- i.e. imposing the restriction that $Cov_4(person, establ.) = \rho_1 Var_4(person)^{1/2} \times Var_4(establ.)^{1/2}$ where subscript 4 refers to the interval 4 value of the statistic and and ρ_1 is the correlation between the person and establishment effects in interval 1. Counterfactual 2 assumes that the variance of establishment effects remains at its interval 1 level. Counterfactual 3 imposes both of these restrictions.

Appendix Table 5: Estimation Results for AKM Model for Full Time Female Workers

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Person and Establishment Parameters:				
Number person effects	9,660,968	10,155,014	9,756,379	9,559,738
Number establishment effects	1,079,129	1,176,133	1,191,607	1,196,201
Summary of Parameter Estimates:				
Std. dev. of person effects (across person-year obs.)	0.332	0.332	0.365	0.397
Std. dev. of establ. effects (across person-year obs.)	0.232	0.227	0.247	0.277
Std. dev. of Xb (across person-year obs.)	0.145	0.098	0.086	0.087
Correlation of person/establ. effects (across person-year obs.)	-0.009	0.039	0.044	0.069
Correlation of person effects/Xb (across person-year obs.)	-0.249	-0.140	-0.117	-0.089
Correlation of establ. effects/Xb (across person-year obs.)	0.040	0.024	0.009	0.041
RMSE of AKM residual	0.137	0.135	0.147	0.157
Adjusted R-squared	0.894	0.900	0.901	0.909
Comparison Match Model				
RMSE of Match model	0.117	0.118	0.125	0.133
Adjusted R-squared	0.922	0.924	0.928	0.934
Std. Dev. of Match Effect*	0.070	0.067	0.077	0.083
Addendum				
Std. Dev. Log Wages	0.420	0.427	0.467	0.521
Sample size	40,846,416	44,351,293	41,576,298	44,751,361

Notes: see notes to Table 4. Models reported here are estimated for subsample of full time female workers. Xb includes year dummies and quadratic and cubic terms in age (total of 8 parameters in intervals 1-3, 9 in interval 4). Match model includes Xb and separate dummy for each job (person-establishment pair).

*Standard deviation of match effect estimated as square root of difference in mean squared errors between AKM model and match effect model.

Appendix Table 6: Decomposition of the Rise in Wage Inequality -- Full Time Females

	Interval 1		Interval 4		Change from Interval 1 to 4	
	Var. Component	Share of Total	Var. Component	Share of Total	Var. Component	Share of Total
	(1)	(2)	(3)	(4)	(5)	(6)
Total variance of log wages	0.176	100.0	0.272	100.0	0.095	100
<u>Components of Variance:</u>						
Variance of person effect	0.110	62.6	0.158	58.1	0.048	50
Variance of establ. effect	0.054	30.6	0.077	28.2	0.023	24
Variance of Xb	0.021	11.9	0.008	2.8	-0.013	-14
Variance of residual	0.014	7.8	0.019	6.9	0.005	5
2cov(person, establ.)	-0.001	-0.8	0.015	5.6	0.017	17
2cov(Xb, person+establ.)	-0.021	-12.1	-0.004	-1.5	0.017	18
Counterfactuals for Variance of log wages: *						
1. No rise in correl. of person/estab. effects	0.176		0.255		0.078	82
2. No rise in var. of estab. effect	0.176		0.247		0.070	74
3. Both 1 and 2	0.176		0.232		0.056	59

Notes: See notes to Table 5. Calculations based on estimated AKM models summarized in Appendix Table 5.

*Counterfactual 1 computes the counterfactual rise in variance assuming the correlation between the person and establishment effects remains at its interval 1 value -- i.e. imposing the restriction that $Cov_4(\text{person, establ.}) = \rho_1 \, Var_4(\text{person})^{1/2} \times Var_4(\text{establ.})^{1/2}$ where the 4 subscript refers to the interval 4 value of the statistic and ρ_1 is the correlation between the person and establishment effects in interval 1. Counterfactual 2 assumes that the variance of establishment effects remains at its interval 1 level. Counterfactual 3 imposes both of these restrictions.