MIT Graduate Labor Economics 14.662 Spring 2015
Lecture Note 1: Wage Density Decompositions

David H. Autor
MIT and NBER

January 30, 2015

1 Introduction

Many of the models we will study this semester are competitive—that is, they assume a free market in equilibrium with no distortions. This baseline assumption is an exceedingly strong one. There are numerous institutional (or not fully competitive) forces acting on the labor market: labor unions, wage mandates, product and labor market regulations, taxes and subsidies, and social norms and strictures (as well as market power, though I'm not sure if we want to think of monopsony/monopoly as an institution). These institutions can affect both prices and quantities in the labor market. In these notes, we'll consider the impact of (some) labor market institutions on the distribution of wages.

Two institutions that have received considerable attention in the United States are labor unions and the federal minimum wage. These institutions and their close relatives are likely to be quite important in many countries. This section of the class examines both institutional forces. Most of the leading papers on these topics are (unfortunately) exclusively based on U.S. data, though that is changing.

The question we explore is how does the minimum wage and/or unionization affect the *shape of the wage distribution* including:

- Overall earnings inequality
- Residual earnings inequality
- Inequality between observably different demographic groups:

The economic importance of this set of topics was not fully appreciated prior to the work of DiNardo, Fortin and Lemieux (DFL, 1996). A glance at the plots of the distribution of wages by gender from their paper—particularly the piling up of women's' wages at the minimum wage threshold show in Figure 1b—makes the case that the minimum wage *must be* (or must have been) important for the wage distribution, particularly for women. But developing a counterfactual is intrinsically difficult. The minimum wage–and similarly labor unions, their other major concern–could affect the wage distribution by:

1. Causing low wage workers to exit the labor force.
2. Boosting wages of workers who earned below the minimum to the level of the minimum
3. Inducing wages in the 'uncovered' sector to either rise or to fall (self-employed workers, tipped workers, illegal immigrants)
4. Causing spillover effects on the wages of workers who are close substitutes to minimum wage workers

5. Causing spill-overs higher up in the overall wage distribution

An over-arching problem in this literature is one of developing counterfactual wage densities, that is an estimate of the entire *distribution* of earnings under another minimum wage or union regime. There are few experiments to work with here, and, moreover this is a general equilibrium problem. So, identification will be tough. A related conceptual problem is whether we should to treat the minimum wage or union penetration as an exogenous forces, or alternatively as equilibrium outcomes of market forces and political economy. For the current discussion, we'll treat the institutions as exogenous while we try to measure their effects on wage outcomes. This is quite imperfect, but it's the right place to start.

In this set of lectures, we will study both the substantive topic of how institutions affect the wage structure and the methodological topic of how to decompose wage densities.

2 Decomposing Wage Distributions: The Challenge

2.1 Kernel density estimation

To study the shape of wage distributions, we need to plot them. Real world empirical distributions are messier than theoretical distribution functions, so we generally need to smooth them to work with them. The main tool for smoothing is the kernel density estimator. A kernel density estimator is an empirical analog to a probability density function. The exact empirical analog of the PDF is simply the conventional histogram. What the kernel density estimator adds to the histogram is smoothing. A kernel density estimate of a distribution of earnings might be calculated as follows:

$$\widehat{f}_h(w) = \sum_{i=1}^{n} \frac{\phi_i}{h} K\left(\frac{w - w_i}{h}\right),$$

where the w_i are wage observations, the $\phi_i ... \phi_n$ are weights with $\sum_i \phi_i = 1$, and $K(\cdot)$ is a kernel density function, normally a Gaussian distribution, and h is the chosen bandwidth, which is effectively the interval of smoothing around any observation. The kernel estimator smooths over nearby neighbors of each observation w_i to estimate the density of w at each given point. The kernel function (normally) accords greater weight to the nearest neighbors of w_i and less weight to more distant neighbors.[1]

[1] This information is likely familiar to many economics students in 2015. When DFL wrote their paper, these techniques were almost unheard of outside select circles.

2.2 An OLS Review[2]

Now that we've plotted the wage distribution and observed that its shape changes over time, we now want to ask what factors explain those changes. That is, we want to decompose changes in the wage distribution into their contributory components. This presents both a statistical and economic challenge. Let's start with the statistical challenge.

Let's start from the basics: the OLS regression. There are several ways to interpret the coefficients from an OLS regression. Those of you who have taken 14.387 know Josh Angrist's favorite: OLS is the best linear approximation of the conditional expectation function. That means the coefficient vector β solves the following minimization.

$$\beta = \arg\min_b \mathbb{E}\left\{ \left(\mathbb{E}\left[Y_i \mid X_i \right] - X_i' b \right)^2 \right\}$$

Consequently, if we want to estimate the OLS regression

$$Y_i = X_i'\beta + \epsilon_i$$

we can recover the exact same coefficients β by using $\mathbb{E}\left[Y_i \mid X_i \right]$ as the dependent variable.

$$\mathbb{E}\left[Y_i \mid X_i \right] = X_i'\beta + \epsilon_i$$

Of course, that's just one interpretation of OLS. Another is that the coefficients β capture the effect of a change in the mean of X on the *unconditional* mean of Y. This result follows from the Law of Iterated Expectations and the CEF result.

$$\begin{aligned} \mathbb{E}\left[Y_i \right] &= \mathbb{E}\left\{ \mathbb{E}\left[Y_i \mid X_i \right] \right\} \\ &= \mathbb{E}\left[X_i'\beta + \epsilon_i \right] \\ &= \mathbb{E}\left[X_i' \right] \beta \end{aligned}$$

Formulated in this way, the coefficients β answer questions like, "If the average level of education in the U.S. increased by two years, by how much would average wages change?" Note that the derivation above relies on the linearity of the expectation operator; that's what allows us to pull the β vector outside the expectation in the last line. This is an important property that will not carry over to quantiles of the wage distribution.

[2] These notes draw partly on recitation lecture notes that TA Sally Hudson wrote for 14.662 in spring 2013

2.3 Oaxaca decomposition

The Oaxaca-Blinder (OB) decomposition, independently developed by Oaxaca and Blinder in 1973, is a canonical tool for separating the influences of quantities and prices on an observed mean difference. It's based on the same principles of OLS: linear expectation functions can be used to analyze mean differences.

For example, if we want to decompose the 'gender gap,' we can write the wages of males and females as:

$$W_M = \bar{X}_M \beta_M \tag{1}$$

$$W_F = \bar{X}_F \beta_F \tag{2}$$

where $\bar{X}_M$ are average demographic characteristics of males (age, education, marital status) and β_M is a vector of 'returns' to these characteristics (and similarly for women).

One can rewrite the gender gap in a variety of ways, including:

$$W_M - W_F = (\bar{X}_M - \bar{X}_F) \cdot \frac{1}{2}(\beta_M + \beta_F) + (\beta_M - \beta_F) \cdot \frac{1}{2}(\bar{X}_M + \bar{X}_F), \tag{3}$$

$$W_M - W_F = (\bar{X}_M - \bar{X}_F)\beta_M + (\beta_M - \beta_F)\bar{X}_M + (\bar{X}_F - \bar{X}_M)(\beta_M - \beta_F), \tag{4}$$

$$\text{and } W_M - W_F = (\bar{X}_M - \bar{X}_F)\beta_F + (\beta_M - \beta_F)\bar{X}_F + (\bar{X}_M - \bar{X}_F)(\beta_M - \beta_F). \tag{5}$$

Each of these decompositions, divides the observed gender wage gap into a component due to differences in quantities between the genders (the $X's$) and differences in 'returns' between the genders (the $\beta's$) and, in the latter two equations, an 'interaction' term.

There are three important things to observe about this decomposition:

1. The Oaxaca technique is useful for decomposing mean differences. It is not applicable to decomposing densities or differences between densities as formulated above.

2. It is a *sequential* decomposition, meaning that the order of the decomposition (or more specifically, the choice of weights) will affect the conclusions. This point is easiest to see by drawing a figure of mean wages by gender against education where $X_M > X_F$ and $\beta_M > \beta_F$. The conclusion about what share of the gap is due to differences in mean education and what share is due to differences in returns will depend on which of the three decompositions above is applied. There is no 'correct' answer, and the choice is generally not innocuous.

3. This technique is intrinsically a partial equilibrium approach. The implicit question the Oaxaca decomposition answers is: what would the gap have been if (a) quantities

did not differ between the genders but prices did or (b) prices did not differ but quantities did? These questions implicitly assume that changing the quantity available of something (education in this example) has no effect on its price (that is, the $\beta's$ are independent of the $X's$). Whether this is a reasonable assumption (or even a useful first order approximation) will depend upon the setting. I would argue that for most of the applications considered below, this is *not* a reasonable assumption.

See also the very nice 2013 AER P&P paper by Patrick Kline, "Oaxaca-Blinder as a Reweighting Estimator," which shows that OB estimator of counterfactual means can be understood as a propensity score reweighting estimator based upon a linear model for the conditional odds of being treated. It's impressive that four decades after OB was introduced and fully assimilated into the standard empirical toolkit, Kline found something genuinely novel to say about it!

2.4 Conditional Quantile Regression Estimation

One natural workaround for wage density decomposition that *does not work* is conditional quantile regression. Quantile regression methods were developed to answer questions about effects of covariates on other features of the outcome distribution aside from the mean. As in OLS, quantile regression coefficients can be modeled as a linear approximation to the conditional quantile function. If we let $\tau \in (0,1)$ denote the τ^{th} quantile of the distribution of log wages given the vector of covariates, then the linear model is

$$Q_\tau (Y_i \mid X_i) = X_i'\beta_\tau + \epsilon_i$$

Note that we estimate a separate coefficient vector β_τ for each quantile τ. This vector solves

$$\beta_\tau = \arg\min_b \mathbb{E}\left\{\rho_\tau (Y_i - X_i'b)\right\}$$

where

$$\rho_\tau (Y_i - X_i'b) = \begin{cases} \tau (Y_i - X_i'b) & \text{for } Y_i - X_i'\beta \geq 0 \\ (1-\tau)(Y_i - X_i'b) & \text{for } Y - X_i'\beta < 0 \end{cases}$$

The function ρ_τ is referred to as the "check function" because the weights will be shaped like a check with the inflection point at $Y_i - X_i'\beta = 0$. (More on this below). Note another key difference between the OLS and CQR minimands: the CQR minimand averages absolute deviations rather than squared deviations. That's why it delivers an estimate that's not sensitive to outliers. You don't reduce the minimand by moving $\hat{\beta}_\tau$ towards the extremes because you pay with deviations from the opposite end.

Please also be clear about the correct interpretation of β_τ . Let's say that X is a dummy variable indicating college versus high-school attendance and the outcome variable is earnings. You estimate a quantile regression for $\beta_{0.5}$. The correct interpretation of $\hat{\beta}_{0.5}$ is generally *not* that it is the effect of college attendance on the median wage earner. Rather, it is the effect of college attendance on the median of the wage distribution. The former interpretation requires an additional rank invariance assumption: that the 50^{th} percentile college earner would also have been the 50^{th} percentile high school earner, so that the treatment (college) preserve rank orderings. This is a very strong and generally untestable assumption.

Unlike the conditional expectation function, the conditional quantile function is not a linear operator. That means that $Q_\tau(Y_i|X_i) = X_i'\beta_\tau$ *does not imply* that $Q_\tau(Y_i) = Q_\tau(X_i)'\beta_\tau$. Consequently, we can't use CQR to answer questions like, "What is the effect of sending more women to college on the 10th percentile of the wage distribution?" Instead, CQR estimates the effect of the college education on the 10th percentile of the wage distribution for a given set of covariates – e.g. young women. That's unfortunate because we typically care about how an economic variable affects the distribution of wages, not the distribution of wages conditional on X. Thus, if the objective is wage density decomposition, the CQR is not the most straightforward place to start.

3 DiNardo, Fortin and Lemieux (1996): Labor Market Institutions and the Distribution of Wages

DFL's 1996 paper proposed a generalization of the OB decomposition that is suited to estimating the impact of economic factors on the shape of the wage distributions. any alternatives have been proposed since their 1996 paper, including the approach developed in Firpo, Fortin and Lemieux's 2011 *Econometrica* paper (and see also their 2013 Handbook of Labor Economics chapter on wage density decompositions). The 1996 paper arguably offers the most straightforward, transparent, and economically lucid attack on the problem.

As above, the Oaxaca decomposition was developed to analyze counterfactual differences in mean earnings. DFL's paper is about estimating and analyzing counterfactual earnings *distributions*. Their technique, a generalization of the Oaxaca distribution, does exactly this. The DFL paper has been influential for three reasons:

1. It first called attention to the possible importance of the minimum wage for the shape of the earnings distribution.

2. It is methodologically elegant, and the methodology can be applied in other settings.

3. The methodology also gets 'extra credit' for cleverness. (It's nearly impossible to overstate the value that economists ascribe to cleverness. Like most obsessions, this one is not altogether healthy.)

3.1 Conceptualization

- DFL view each wage observation in a given distribution as a vector composed of the wage itself, a set of individual attributes, z, and a time subscript t_w. Thus, they write $w_i \equiv (w_i, z, t)$.

- They express the observed distribution of wages at time $f_t(w)$ as a joint distribution of wages and attributes conditional on time $f(w, z|t_w; m_t)$, integrated over the distribution of individual attributes $F(z|t_z)$ at date t_z. Under this notation, m_t refers to the minimum wage prevailing at time t.

- Putting these pieces together, DFL write the joint distribution of wages and attributes, conditioning on time t as:

$$f_t(w) = \int_{z \in \Omega_z} dF(w, z|t_{w,z} = t; m_t). \tag{6}$$

  This expression describes the marginal density of wages at t by integrating out the distribution of attributes z, and conditioning on the time period and minimum wage level. Note that Ω_z is the domain of individual attributes and t for purposes of their paper will correspond to one of two points in time, 1979 and 1988.

- Rewriting (6) and applying the law of iterated expectations gives:

$$f_t(w) = \int_{z \in \Omega_z} f(w|z, t_w = t; m_t) dF(z|t_z = t) \equiv f(w; t_w = t, t_z = t, m_t).$$

- Here, the distribution of w is expressed conditional on z and the distribution of z is expressed conditional on t. Iterating expectations allows us to condition on additional variables such as z and then integrate them out again. Nothing is gained (so far) from this, but the value of conditioning will be apparent in a moment.

- Attending to notation here is important. $f(w; t_w = 88, t_z = 88, m_{88})$ refers to the observed distribution of wages in 1988, that is the wage distribution conditional on the distribution of $z's$ prevailing in 1988, the level of the minimum wage in 1988, and the prices associated with characteristics z in 1988 ($t_w = 88$ refers to the price level). Thus, DFL are conceptualizing the wage distribution as being composed of three primitives:

1. The level of the minimum wage, m_t. This is the *institution* parameter.
2. The distribution of worker attributes, t_z. This is the *quantity* matrix.
3. The conditional wage distribution for given attributes, t_w. This is the *price* function, and it is quite central. The price function is in effect a distribution function that "links" $z's$ to wages. This function is the conditional distribution of wages for each unique entry in the z matrix. Integrating the price function over the distribution of $z's$ (the quantity matrix) produces the observed wage distribution (ignoring for a moment m_t).

- Using this notation, the expression $f(w; t_w = 88, t_z = 79, m_{88})$ refers to the *counterfactual* distribution of wages in 1988 with the distribution of $z's$ at its 1979 level, but prices and minimum wages at their 1988 levels.

- Under the *strong* assumption that the 1988 structure of wages does not depend on the distribution of $z's$, the hypothetical density $f(w; t_w = 88, t_z = 79, m_{88})$ can be written as

$$\begin{aligned} f(w; t_w = 88, t_z = 79, m_{88}) &= \int f(w|z, t_w = 88; m_{88}) dF(z|t_z = 79) \\ &= \int f(w|z, t_w = 88; m_{88}) \psi_z(z) dF(z|t_z = 88), \end{aligned} \tag{7}$$

  where the 'reweighting' function $\psi_z(z)$ maps the 1979 distribution of $z's$ into the 1988 distribution.

- The reweighting function is defined as

$$\psi_z(z) = \frac{dF(z|t_z = 79)}{dF(z|t_z = 88)}. \tag{8}$$

- This expression is simply the ratio of probability mass at each point z in the year 1979 relative to 1988. Hence, $\psi_z(z)$ reweights the 1988 density so that observations that were relatively more likely in 1979 than 1988 are weighted up and observations that are relatively less likely are weighted down.

- Notice that this reweighting is non-parametric (so far); we are not specifying the function that maps $z's$ into $w's$. We are simply reweighting the observed joint density of $z's$ and $w's$.

3.2 The clever part

- As a conceptual matter, $\psi_z(z)$ is straightforward to implement. As an empirical matter, it is not.

- Notice that $\psi_z(z) = dF(z|t_z = 79)/dF(z|t_z = 88)$ is a high dimensional object to estimate because z exists in $\Re^k$ where k is the number of items in the vector of characteristics (e.g., age, gender, education, race, union status, etc.).

- So, estimating ψ_z in (8) would be fraught with imprecision; in finite data, there will likely be probability mass zero in the numerator or denominator for some values of z. DFL solve this implementation problem by a savvy application of Bayes rule.

- A quick reminder on Bayes' rule:

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\sum_i \Pr(B|Z_i) \cdot \Pr(Z_i)}. \tag{9}$$

- Bayes' rule calculates the posterior probability of event A given an observation of B and *prior* knowledge of the joint distribution of B and other events Z (one of which is A). Intuitively, we are given some event B and we want to know give how likely it is that some other event A occurred given B. The answer is simply the ratio of the probability of the joint occurrence of A and B to the total occurrences of B, which is the expression above.

- Applying this method to the DFL problem, we get:

$$\Pr(z|t_z = 79) = \frac{\Pr(t_z = 79|z) \cdot dF(z)}{\int_z \Pr(t_z = 79|z) \cdot dF(z)}, \tag{10}$$

  and

$$\Pr(z|t_z = 88) = \frac{\Pr(t_z = 88|z) \cdot dF(z)}{\int_z \Pr(t_z = 88|z) \cdot dF(z)}. \tag{11}$$

  Hence,

$$\widehat{\psi}_z = \frac{\Pr(t_z = 79|z)}{\Pr(t_z = 88|z)} \cdot \frac{\Pr(t_z = 88)}{\Pr(t_z = 79)}. \tag{12}$$

- Unlike (8), equation (12) can be readily estimated. There are three steps:

  1. Pool data from both years of the sample, 1979 and 1988.
  2. Estimate a probit model for

$$\Pr(t_z = t|z) = \Pr(\varepsilon > -\beta H(z)) = 1 - \Phi(-\beta H(z)),$$

which is simply the likelihood that an observation is from year t given z.

3. Use estimates of $1 - \Phi(-\beta H(z))$ to form $\widehat{\psi}_z$ for each observation.

- The subtlety here is that DFL are conceptualizing time as part of the state space.

- By estimating time as a function of the $z's$ (i.e., whether a given observation i is from 1988 or 1979 based on its $z's$), DFL use Bayes' rule to invert the high dimensional problem in (6) into a unidimensional problem.

- By the way, this reweighting transformation using the probit model is parametric (that is, the probit model assumes a functional form). The entire DFL method is therefore called 'semi-parametric' because of the parametric reweighting approach.

- Of course, one can always make a 'non-parametric' by using a flexible functional form and 'promising' to make it more flexible if more data were to arrive.

- As you have observed, there is an important underlying assumption that DFL underscore in the text:

  > "Calling the counterfactual density $f(w; t_w = 88, t_z = 79, m_{88})$ the 'density that would have prevailed if individual attributes had remained at their 1979 level' is a misuse of language. This density should rather be called the "density that would have prevailed if individual attributes had remained at their 1979 level *and* workers had been paid according to the wage schedule observed in 1988," since we ignore the impact of changes in the distribution of z on the structure of wages in general equilibrium."

- Hence, the DFL approach is the exact analog of the Oaxaca approach. It holds prices at their 1988 levels while imposing the 1979 distribution of quantities to simulate a counterfactual wage distribution. This allows DFL to account for the impact of changing $z's$ and changing t_z in their decomposition.

3.3 Accounting for unionization and minimum wages

In the case of unionization and minimum wages, there is not a natural analogy to the Oaxaca approach. These outcomes are not (wholly) individual characteristics that are 'priced' by the market. Rather they potentially spillover across the distribution.

To incorporate them into the framework, DFL take different approaches for these two factors. They treat unionization as a wholly *individual* trait. Conversely, they treat the minimum wage as a *distributional* characteristic.

Unions

To treat unionization as an individual characteristic, the following assumptions are needed:

1. There is no non-random selection of union status by ability. Hence, the observed union wage effect from a cross-sectional regression corresponds to the true causal effect.

2. There are no general equilibrium impacts of unionization on the wage distribution for non-members–union-threat effects, employment effects, spillover effects.

These assumptions allow DFL to "Estimate the density of wages that would have prevailed if unionization had remained at its 1979 level and workers were paid according to the union and non-union wage schedules observed in 1988 [or 1979]." In short, unionization is simply treated as a z variable.

Minimum wages

The assumptions required to form a counterfactual minimum wage estimate are (even) more stringent:

1. Minimum wages have no spillover effects on the distribution of wages above the minimum. This is a conservative assumption since any spillover effects (which are plausible slightly higher in the distribution) would *augment* the impact of the minimum wage.

2. The shape of the conditional density of wages at or below the minimum depends only upon the real minimum.

3. The minimum wage has no impact on employment probabilities, hence there is no need to develop counterfactual wage densities for workers who lose employment due to imposition of a binding minimum. This assumption is also conservative since removal of low wage observations from the distribution (due to job loss) would tend to further decrease inequality.

These assumption allow DFL to 'graft' the lower tail of the earnings distribution below the minimum wage (e.g., from 1979) directly onto another era's wage distribution (e.g., 1989) when imposing the counterfactual minimum wage. This is not entirely satisfactory, but it is difficult to improve upon in this setting. The net results are visible in DFL Figure 3. Note that the apparent spill-overs just to the right of the minimum wage in the counterfactual density (Panel C) are an artifact of the kernel density estimator which smooths over the discontinuity.

3.3.1 Accounting for supply and demand

Supply and demand is notably absent from the analysis so far. DFL do the following:

- Create 32 education-experience-gender cells n_j
- Calculate changes in log relative supply in each $\Delta\widehat{n}_j$
- Estimate changes in log relative demand within each cell $\Delta\widehat{d}_j$ using a 'fixed coefficients manpower requirements index reflecting between sector shifts in relative labor demand' (this is essentially an industry shift measure).
- Calculate the implied changes in wages $\Delta\widehat{w}_j$ in each cell based assuming that $\sigma = 1.7$. (Note that the CES form implies that the elasticity of substitution among *all* cells is equal to σ. We will discuss the CES model at greater length in the weeks immediately ahead.)

Hence, in DFL's notation:

$$f(w|z, t_{w=88}; m_{88}; d_{79}; n_{79}) = f(w - \Delta\widehat{w}_j|z, t_{w=88}; m_{88}; d_{88}; n_{88}). \tag{13}$$

- DFL is the only paper in this entire wage-density decomposition literature that takes supply and demand seriously. This is much to the paper's credit. Subsequent authors—including these same authors at later dates—have apparently concluded that supply and demand do *not* need to be taken into account when constructing counterfactual wage structures. Even to a reduced form empiricist, that conclusion might seem a bit startling.

3.4 Results

The easiest way to see the results of the DFL analysis is to study Figures 6 and 7. Rather than giving the actual and counterfactual PDFs, these figures plot the *differences* between the 1979 and 1988 densities (note: the integral of each figure must therefore be zero). The successive panels of the figure show the remaining differences in densities after applying each of the counterfactual calculations: the minimum wage, unionization, individual attributes, and supply and demand. If the final density difference was completely flat and lay on the x-axis, it would indicate that the model explained all observed changes in wage distributions.

Three observations:

- After applying these counterfactuals, much growth in inequality remains, particularly in the upper tail of the distribution (this is seen by the extra mass above the x-axis at the right hand tail and the missing mass in the middle).

- The exercise appears to have much more explanatory power for earnings inequality among women than men. This is not surprising because the minimum wage had been much more binding for women who compose a disproportionate share of the lower tail of the wage distribution.

- Because the decomposition is 'sequential,' the choice of the ordering of factors is *not* innocuous. For example, by placing supply and demand last in the decomposition, DFL give it less weight. By comparing Tables III and V of their paper, you can see that the importance of some factors more than doubles if their contributions are calculated first rather than last. As above, no sequence is 'correct.' DFL's choice of sequence is probably not accidental.

Some key results of DFL for Males:

	Measure	Change	MinWg	Unions	X's	S&D	Unexplained
MinWg 1st	90-10	0.195	25.3%	10.7%	20.7%	20.7%	22.6%
S&D 1st			16.0%	24.1%	4.8%	32.6%	
MinWg 1st	50-10	0.076	65.7%	-25.6%	49.7%	10.9%	-0.7%
S&D 1st			43.5%	-5.5%	26.9%	33.9%	

Key results of DFL for females:

	Measure	Change	MinWg	Unions	X's	S&D	Unexplained
MinWg 1st	90-10	0.328	45.1%	1.3%	25.6%	11.1%	16.9%
S&D 1st			41.6%	0%	14.6%	26.8%	
MinWg 1st	50-10	0.243	61.7%	-4.1%	32.1%	-4.5%	14.8%
S&D 1st			56.5%	-3.0%	18.7%	13.0%	

A fair reading of these results suggests three conclusions:

- The decline in the U.S. minimum wage has probably been quite important to the growth of earnings inequality in the lower tail of the distribution for females.

- The impacts of minimum wages on male earnings inequality are less significant, though not trivial at the bottom of the distribution.

- For males, supply and demand are likely to have been quite important to the growth of earnings inequality.

How convincing is this analysis? Many would criticize this approach for applying a partial equilibrium analysis to a general equilibrium problem. In addition, an open question for this type of analysis is the conceptual problem of how economists should view the exogeneity/endogeneity of labor market institutions. (The DFL approach takes them as exogenous.)

Despite these limitations—and all papers have some—the DFL approach was original, insightful and productive in causing economists to take the minimum wage more seriously as an explanation for rising wage inequality. Several years later, David Lee's (1999) paper offered a potentially more compelling empirical case for its importance. We won't have time to discuss Lee's paper in class, but you may find it instructive to read. In that case, you might also read the working paper by Autor, Manning and Smith (2010) which offers something of a critique and reanalysis of Lee.

4 Decomposing wage densities: Some techniques and results

The DiNardo, Fortin and Lemieux 1996 paper presents one clever approach to decomposing wage densities. But there are a number of alternatives. I'll review them here, but I won't spend class time on them.

4.1 The Juhn, Murphy and Pierce (1993) decomposition

JMP's highly cited 1993 paper set out to summarize the rising dispersion of earnings in the U.S. during the 1970s and 1980s. JMP wanted a tool for describing the components of wage density changes that could be attributed to measured prices, measured quantities and residuals (which they referred to as unmeasured prices and quantities). Figure 1 of their paper reminds you of the dramatic wage structure changes that motivate their work.

The technique that JMP developed, the "full distribution accounting method" features both good ideas and important limitations. This technique has been widely applied since the publication of their paper and so is worth spending a few minutes on.

The wage equation in time t can be written as:

$$Y_{it} = X_{it}\beta_t + u_{it}. \tag{14}$$

Write u_{it} as,

$$u_{it} = F^{-1}\left(\theta_{it}|X_{it}\right). \tag{15}$$

where $F^{-1}\left(\cdot|X_{it}\right)$ is the inverse cumulative distribution of wage residuals conditional on X_{it}, and θ_{it} is $i's$ percentile rank in the residual distribution. Notice that both θ_{it} and $F\left(\cdot|X_{it}\right)$ will depend upon the conditioning variables X.

In this framework, changes in inequality over time come from three sources: (1) changes in the distribution of observable individual characteristics, X, (2) changes in the returns to those observable characteristics, β; and (3) changes in the distribution of residuals $F(\theta|X)$.

If we define $\bar{\beta}$ as the average price of observables over some time interval and $\bar{F}(\cdot|X_{it})$ as the average cumulative distribution of residuals, then we can write the *difference* in inequality between the current period and the 'average' periods as:

$$Y_{it} = X_{it}\bar{\beta} + X_{it}\left(\beta_t - \bar{\beta}\right) + \bar{F}^{-1}(\theta_{it}|X_{it}) + \left[F_t^{-1}(\theta_{it}|X_{it}) - \bar{F}^{-1}(\theta_{it}|X_{it})\right]. \qquad (16)$$

This equation potentially allows us to simulate counterfactual distributions by varying prices, quantities or the residual distribution.

For example, if we wanted to calculate the counterfactual distribution of wages holding fixed observable prices and the residual distribution at their averages and varying only the distribution of $X's$, we could calculate:

$$Y_{it}(1) = X_{it}\bar{\beta} + \bar{F}^{-1}(\theta_{it}|X_{it}). \qquad (17)$$

If we want to allow observable quantities and prices to vary with time, we can write:

$$Y_{it}(2) = X_{it}\beta_t + \bar{F}^{-1}(\theta_{it}|X_{it}). \qquad (18)$$

Finally, if we want to allow quantities, observable prices and unobservables to move simultaneously, we end up with:

$$Y_{it}(3) = X_{it}\beta_t + F^{-1}(\theta_{it}|X_{it}) = Y_{it}. \qquad (19)$$

Using this accounting scheme, JMP propose the following decomposition:

1. $Y_{it}(1) - \bar{Y}_i$ is the component of the difference in inequality between t and the average period due to changing quantities.

2. $Y_{it}(2) - \left(Y_{it}(1) - \bar{Y}_i\right)$ is the marginal contribution of changing prices..

3. $Y_{it}(3) - Y_{it}(2)$ is the marginal contribution of changing residuals.

Notice that

$$[Y_{it}(3) - Y_{it}(2)] + \left[Y_{it}(2) - \left[Y_{it}(1) - \bar{Y}_i\right]\right] + \left[Y_{it}(1) - \bar{Y}_i\right] = Y_{it}(3) = Y_{it} \qquad (20)$$

so adding the pieces together recovers the actual distribution of wages in period t. This is of course an identity.

More specifically, JMP perform their estimation in there steps:

1. To obtain the effect of quantities only, JMP predict wages for all workers in the sample in year t using the average coefficient vector $\bar{\beta}$, and computing a residual for each worker based on his rank in year $t's$ residual wage distribution and applying the average cumulative residual distribution over the full sample.

2. To obtain the marginal effect of quantities, they repeat this procedure, now using the coefficient vector β_t but retaining the average residual distribution.

3. Finally, the third component is simply calculated as the difference between the actual wage distribution in year t and the counterfactual in step (2). Thus, notice that the residual component is itself calculated as a residual. As discussed below, this reflects a shortcoming of the procedure: the subcomponents don't necessarily add to the whole (despite equation (20)).

Like the Oaxaca-Blinder and DFL's decompositions, the JMP decomposition is sequential. Depending on the order of the decomposition, one will generally attribute different shares to quantities, prices, and residuals. For example, Goldin and Margo (1992 *QJE)* find that the sequence of the decomposition is substantively important for decomposing the components of changing inequality in the 1940s.

A key virtue of the JMP decomposition is that can be applied to any wage quantity (i.e., the $10^{th}, 50^{th}$, or 90^{th}) percentile. Hence, like DFL – and unlike Oaxaca-Blinder – the JMP decomposition can ostensibly be used to simulate entire counterfactual distributions.

An interesting conceptual/rhetorical point here is that JMP are purportedly decomposing the residual into two pieces: a 'location' parameter and a 'scale' parameter – that is, a residual rank, and corresponding residual wage. This is clever but is only strictly correct under unappealing assumptions, specifically that residuals are exclusively accounted for by unmeasured prices and quantities, rather than measurement error and luck, and moreover, that a person's rank in the residual distribution is invariant to the set of conditioning $X's$. Without this strong 'single index' interpretation of the wage residual, a person's rank in the residual distribution, θ_{it}, is not necessarily indicative of their 'unmeasured quantity of skill' (with $F^{-1}(\theta_{it}|X_{it})$ giving its price).

The JMP methodology has two main shortcomings. The first is that the OLS wage regression at the heart of the JMP technique provides a model for the conditional mean of the wage distribution and its results do not extend naturally to wage quantiles. For example, there is no presumption that $F_t(50) = 0$, and so it need not be the case that

$\bar{y}_t = \bar{X}_t' \beta_t + F_t(50)$. However, it will be the case that $\bar{y}_t = \bar{X}_t' \beta_t + E[F_t(\cdot)]$. So, in practice, this inconsistency may not be substantively important (though it's certainly inelegant).

A more significant shortcoming is that the components of the JMP counterfactual decomposition need not sum to the total observed change. Specifically, adjusting the wage distribution in period t for the change in each of the three accounting components (prices, quantities, residuals) between periods t and τ will not generally recover the observed wage distribution in period τ. The reason is that knowledge of the marginal distributions of $X'\beta$ and $F(\theta)$ is not generally sufficient to characterize the marginal distribution of $w = X'\beta + F(\theta)$. Since $X'\beta$ and $F(\theta)$ are random variables, the distribution of their sum $g(w = X'\beta + F(\theta))$, depends on both their variances and covariance (i.e., their joint distribution). This covariance is not accounted for by the JMP decomposition unless $F(\theta)$ is allowed to depend explicitly on X.

Does it? In the paper, JMP denote $F_t^{-1}(\theta)$ as depending explicitly on X (e.g., $F^{-1}(\theta|X)$). But, as noted by Lemieux 2002, it is unclear how the dependency of the residual distribution on X is implemented in practice. Because X contains continuous covariates, there are essentially an unlimited number of conditional distributions of the residual. In most subsequent implementations of the JMP decomposition (see especially Blau and Khan, 1994), researchers have used an unconditional residual distribution, $F_t^{-1}(\theta)$.

There is one special case, however, where the marginal distribution of w is recoverable from the marginal distributions of $X'\beta$ and $F(\theta)$. This is when these marginal distributions are independent. Under classical OLS assumptions—in particular, a normally distributed, homoskedastic error term—this independence condition will hold. But this is extremely restrictive. In practice, residual wage dispersion does appear to vary significantly with education and experience, a point made forcefully by Lemieux's 2006 *AER P&P* paper.

The key results of the JMP decomposition are found in Table 4 of the paper. For the period of immediate interest in their paper, 1979 - 1988:

1. About 40 percent of the growth of 90/10 inequality occurs above the median (90/50) and 60 percent below the median (50/10).

2. Changes in quantities ($X's$) play almost no role in the growth of inequality.

3. But changes in prices do. Observed prices play an equally large role above and below the median, and account for about 55 percent of the total growth in inequality.

4. The residual ('unobserved quantities and prices') accounts for the other 45 percent. Of that component, 75 percent of the growth in the residual is found below the median.

Another interesting finding is that while the 'observed price' components do not contribute to rising inequality until the 1980s, residual inequality starts to rise sharply in the 1970s (see Figure 7). But, as has been later demonstrated by other researchers, this difference in the timing of residual versus between-group inequality is not present in all data sets. The CPS May/ORG data do not show strong growth in residual inequality until the 1980s. The CPS March data, which JMP use, do. It also appears from subsequent analysis that JMP overstate the magnitude of the rise in residual inequality in the March during the 1970s.

The uncertainty about the basic facts surrounding the timing of residual inequality has generated significant controversy in the U.S. inequality literature. Some see the purported rise in residual inequality in the 1970s as evidence of some form of SBTC in this decade. Those who believe that residual inequality started rising in the 1980s, coincident with the rise in between-group inequality, tend to believe that all of the rise in inequality is due to a single phenomenon. This controversy remains active. I will not devote class time to it.

4.2 Comparison of JMP and DFL

The JMP tool does two things that DFL does not do:

First, JMP explicitly models the role of 'residual prices and quantities.' In contrast, DFL's wage density technique does not distinguish between between-group and residual prices. However, Lemieux 2006 in the *AER* proposes a simple extension whereby the DFL technique is applied to the residual wage distribution (estimated using an OLS model) after the between group effects have been purged via OLS.

Alongside these virtues, the JMP approach has two weaknesses. First, it does not 'add up' – the sum of components need not equal the total change. And moreover, JMP is conceptually not very clean because it uses an OLS regression, a model of the conditional mean, to do quantile analysis. (The DFL model is also a strange hybrid of parametric and non-parametric tools, i.e.., the logit or probit reweighting equation married to the non-parametric wage distribution). And this conceptual flaw is exacerbated if one additionally uses OLS regressions to generate the wage residuals for the DFL residual reweighting (as proposed by Lemieux 2006).

It would be desirable to have a unified model that features the strengths of JMP (explicit decomposition of wage distribution into quantities, between-group prices and residuals) and the strengths of DFL (adds up; conceptually almost-kosher). One such approach is described below.

5 Wage density decomposition: A quantile regression approach

We'll now consider an approach to wage density decomposition based on quantile regressions. This approach was proposed by Machado and Mata (*Journal of Applied Econometrics*, 2005) and slightly extended by Autor, Katz and Kearney (2005, NBER Working Paper).

5.1 Quantile regression basics

Let $Q_\theta(w|X)$ for $\theta \in (0,1)$ denote the θ^{th} quantile of the distribution of the log wage given the vector of covariates. We model these conditional quantiles as

$$Q_\theta(w|X) = X'\beta(\theta), \tag{21}$$

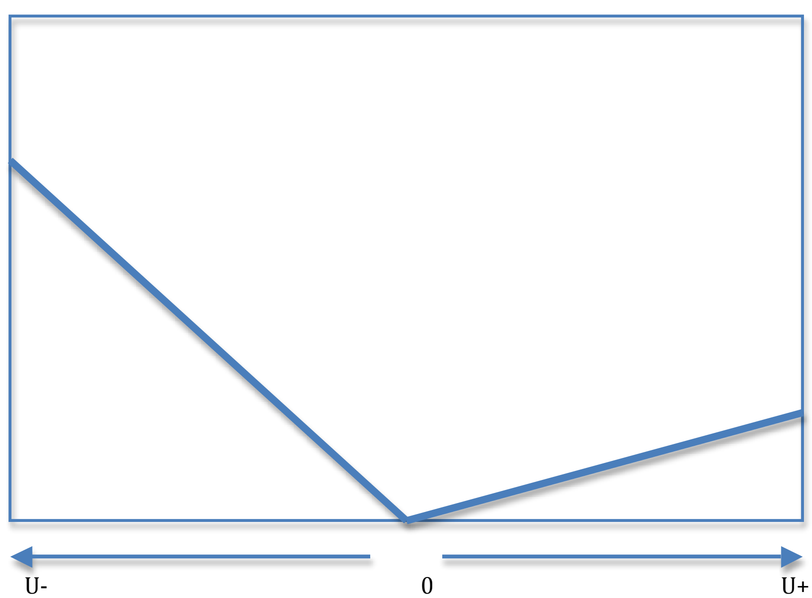
where X is a $k \times 1$ vector of covariates and $\beta(\theta)$ is a conformable vector of quantile regression (QR) coefficients. For given $\theta \in (0,1)$, $\beta(\theta)$ can be estimated by minimizing in β,

$$n^{-1}\sum_{1=1}^{n} \rho_\theta(w_i - X_i'\beta) \tag{22}$$

with

$$\rho_\theta(\mu) = \begin{cases} \theta\mu & \text{for } \mu \geq 0 \\ (\theta - 1)\mu & \text{for } \mu < 0 \end{cases}. \tag{23}$$

The latter expression is referred to as the "check function" because the weight applied to μ will be shaped like a 'check' with the inflection point at $w_i - X_i'\beta = 0$. Check Function for $\theta = 0.25$:



This expression often looks mysterious at first, but it is simple to demonstrate to yourself that it works. For example, consider the sequence $[0, 11, 23, 27, 40, 50, 60]$. What's the median? By inspection, it's 27. Now, plug in above. Note that there are no $X's$ here, we are just estimating the constant in expression (22). So, we have:

$$\begin{aligned} z &= \frac{1}{7} \cdot \left\{ \begin{array}{c} -.5\left[(0-27)+(11-27)+(23-27)+(27-27)\right] \\ +.5\left[(40-27)+(50-27)+(60-27)\right] \end{array} \right\} \\ &= \frac{1}{7} \cdot \left[-.5\left(-27-16-4\right)+.5\left(13+23+33\right)\right] = \frac{1}{7} \cdot 58 \end{aligned}$$

Try adjusting $\hat{\beta}$ upward or downward from 27. You will quickly discover that you cannot do better than $\hat{\beta} = 27$. (Easiest to use a spreadsheet and replace the if...then expression in equation (23) with the absolute value operator, though this only works for the median.)

Now, try something counterintuitive. Add a second zero to the bottom of the sequence and a very large positive number to the top of the sequence (say $10,000$). Intuition might suggest that this would increase $\hat{\beta}$ to > 27. But that is not the case. Because you are minimizing the sum of absolute deviations $\sum \rho_\theta \left(w_i - X_i'\beta\right)$, you do not gain by moving $\hat{\beta}$ towards the extremes; you pay with deviations from the opposite end.

Beyond understanding the linear programming problem, you should be pictorially clear on what a quantile regression is estimating, as well as the correct interpretation of $\beta\left(\theta\right)$. For example, let's say that X is a dummy variable indicating college versus high-school attendance and the outcome variable is earnings. You estimate a quantile regression for $\beta\left(50\right)$. The correct interpretation of $\hat{\beta}\left(\theta\right)$ is generally *not* the effect of college attendance on the 50^{th} percentile wage earner. Rather, it is the effect of college attendance on the 50^{th} percentile of the wage distribution. The former interpretation requires an additional 'rank invariance' assumption: the 50^{th} percentile college earner would also have been the 50^{th} percentile high school earner (i.e., the treatment preserves rank orderings). This is a very strong and generally untestable assumption (notably, it is an assumption in the JMP decomposition). Many users of quantile regressions mistakenly interpret their results using an implicit rank invariance assumption.

5.2 A quantile model of wage inequality: Estimation for given X

As discussed by Machado and Mata, if equation (21) is correctly specified, the conditional quantile process—that is, $Q_\theta\left(w|X\right)$ as a function of $\theta \in (0,1)$—provides a full characterization of the conditional distribution of wages given X. Realizations of w_i given X_i can be viewed as independent draws from the function $X_i'\beta\left(\theta\right)$ where the random variable θ is uniformly distributed on the open interval $\theta \in (0,1)$.

Let us say we have accurately fit the conditional quantile function $Q_\theta(w|X)$ at a sufficiently large number of points θ. We can use the estimated parameters $\hat{\beta}(\theta)$ to simulate the conditional distribution of w given X. Here, we use the Probability Integral Transformation: If U is a uniform random variable on $[0,1]$, then $F^{-1}(U)$ has the density $F(\cdot)$. Thus, if $\theta_1, \theta_2 ... \theta_j$ are drawn from a uniform $(0,1)$ distribution, the corresponding j estimates of the conditional quantiles of wages at X, $\hat{w} \equiv \{X'\beta(\theta_i)\}_{i=1}^{j}$, constitute a random sample from the (estimated) conditional distribution of wages given X.

This simulation procedure characterizes the conditional quantiles of the w for given X. It does not provide the *marginal density* of w. This is because the marginal density depends upon both the conditional quantile function, $\hat{\beta}(\theta)$, and the distribution of the covariates $g(X)$. (For simplicity, we will treat $g(X)$ as known rather than estimated).

Getting the model right may also be important. The conditional quantile model will hold *exactly* in a case where both location and scale depend linearly on the covariates (for example in the classical location shift model where $w = X_i'\beta + \varepsilon$ and ε is a normal, iid error term). In more general cases, the conditional quantile model may provide a reasonable approximation to the true conditional quantile, and this approximation can generally be improved by specifying flexible functions of X when estimating $\beta(\theta)$.

5.3 Estimating the marginal density of w using the quantile model

To generate a random sample from marginal density of w, we can draw rows of data X_i from $g(X)$ and, for each row, draw a random θ_j from the uniform $(0,1)$ distribution. We then form $\hat{w}_i = X_i'\hat{\beta}(\theta_j)$, which is a draw from the wage density implied by the model. By repeating this procedure, we can draw an arbitrarily large random sample from the desired distribution. This procedure – successively drawing from $g(X)$ and θ to form $\hat{w}_i$ – is equivalent to numerically integrating the estimated conditional quantile function $\hat{Q}_\theta(w|X)$ over the distribution of X and to θ form

$$f(\hat{w}) = \int \int_{X,\theta} \hat{Q}_\theta(w|X)\, dX d\theta$$

This decomposition has two useful properties. First, the conditional quantile model partitions the observed wage distribution into 'price' and 'quantity' components. [But it is worthwhile to ask what these 'prices' mean.] This is similar to a standard Oaxaca-Blinder procedure using OLS regression coefficients, with the key difference that the OLS model only characterizes the central tendency of the data (i.e., the conditional mean function, describing 'between-group' inequality). By contrast, the conditional quantile model characterizes both the central tendency of the data (in this case, the median) and the dispersion of the outcome variable conditional on X, i.e., the wage 'residuals.' This feature is critical for estimating the impact of composition on the shape of the residual wage distribution.

Second, under the standard—but unappealing—wage decomposition assumption that aggregate skill supplies do not affect aggregate skill prices, we can use the conditional quantile model to simulate the impact of changing composition or prices on distribution of wages. In particular, by applying the labor force composition data $g_t(X)$ from a given time period to the price matrix $\hat{\beta}_\tau(X)$ from any other time period τ, we can simulate the counterfactual distribution of wages that would prevail if labor force composition were given as in time period t and labor market prices were given as in time period τ. Note that because the $\hat{\beta}_\tau(X)$ matrix describes the conditional distribution of wages for given values of X, this simulation captures the effects of composition on both between-group and residual inequality.

5.4 Extension to residual inequality

While Machado-Mata do not extend their approach to estimating counterfactual measures of *residual* inequality, this extension is straightforward (Autor, Katz, Kearney's 2005 NBER Paper make this extension, as does Melly 2005).

Define the coefficient vector $\hat{\beta}(50)$ as the measure of 'between-group' inequality, denoted as $\hat{\beta}^b \equiv \hat{\beta}(50)$. $\hat{\beta}^b$ serves a role akin to $\hat{\beta}_{OLS}$ in a conventional Oaxaca-Blinder decomposition. In the standard application, $\hat{\beta}_{OLS}$ provides a measure of between-group inequality; it estimates the central tendency of the data conditional on X. In the quantile application, the central tendency measure is the conditional median, estimated by $\hat{\beta}^b$.

Following this logic, define a measure of within-group inequality as the difference between the estimated coefficient vector $\hat{\beta}(\theta)$ and the median coefficient vector $\hat{\beta}^b$:

$$\hat{\beta}^w(\theta) \equiv \left[\hat{\beta}(\theta) - \hat{\beta}^b\right] \text{ for } \theta \in (0,1). \tag{24}$$

By construction, $\hat{\beta}^w(50) = 0$. Hence, the residual quantile coefficient matrix is purged of 'between-group' inequality, and measures the expected dispersion of w at any given value of X, holding the conditional median at zero. By applying the coefficient matrix $\hat{\beta}^w(\theta)$ to the distribution of covariates, $g(X)$, one can calculate the (estimated) dispersion of w that is exclusively attributable to residual inequality. If, for example, $\hat{\beta}(\theta) = \hat{\beta}^b \; \forall \; \theta$, then residual inequality is zero in this model.

To summarize, the (correctly specified) conditional quantile model provides a complete characterization of the distribution of w as a function of three components: the distribution of covariates, $g(X)$, the vector of between-group prices, $\hat{\beta}^b$, and the matrix of within-group (residual) prices $\hat{\beta}^w$. We write $f_t(\hat{w}_t) \equiv f\left(g_t(X), \hat{\beta}_t^b, \hat{\beta}_t^w\right)$

5.5 Implementation

This procedure is implemented in three steps.

1. Estimate quantile coefficient vectors for each time period. For example, AKK use the May/MORG March samples to estimate a standard Mincer wage model fit using quantile rather than OLS regressions. For each sample, year, and gender, they estimate models for quantiles,
$$[0.1, 0.3, ..., 99.7, 99.9]$$
at intervals of one-fifth of a centile, with one additional model fit for the median (quantile 50.0). These QR coefficients, $\hat{\beta}(\theta)$, provide the 'prices' for the simulation exercise. Here, $\hat{\beta}(\theta)$ is a $k \times m$ matrix of quantile regression coefficients, where k is the number of elements in X and m is the number of quantiles (501) estimated in θ.

2. Calculate the residual price vector $\hat{\beta}_t^w$ using equation (24). This yields $\hat{\beta}_t^b$, a $k \times 1$ vector of between-group prices, and $\hat{\beta}_t^b$, a $k \times (m-1)$ matrix of 'within-group prices.'

3. Draw simulated data from the distribution $f\left(g_t(X), \hat{\beta}_t^b, \hat{\beta}_t^w\right)$ by applying the price matrices $\hat{\beta}_t^b$, $\hat{\beta}_t^w$ to the rows of $g_t(X)$.

Before applying the technique to simulate counterfactual distributions, AKK check the performance of the model for replicating observed (actual) distributions of overall and residual inequality. AKK apply the QR coefficients $\hat{\beta}_t^b$, $\hat{\beta}_t^w$ to the quantity series, $g_t(X)$ from the contemporaneous time period to generate simulated wage distributions. The simulated series are shown in Appendix Figure 1 of their paper. If the QR model were a perfect fit to the conditional wage distributions, these series would exactly overlap one another. In practice, the discrepancy between the actual statistic and the model-based simulations is small enough to be undetectable in most cases.

To benchmark the performance of the quantile simulation procedure for *residual* wage inequality, AKK compare simulated and actual 90/10 residual wage dispersion by year and gender (Appendix Figure 2). The series labeled 'Observed Median Reg Residual' is formed from median regressions of log hourly earnings on the covariates above. The series 'Simulated Median Reg Residual' presents the corresponding statistics for the simulated series formed using $\hat{\beta}_t^w$ and $g_t(X)$. Since almost all prior residual decompositions analyze OLS regression residuals, AKK also plot 90/10 residual inequality for an identically specified OLS model.

As is apparent from Appendix Figure 2, OLS and Median Regression residuals have nearly identical dispersion in the AKK application. This indicates that the distinction between mean and median regression is not substantively import for interpreting these residual decomposition results. As is the case for the overall inequality series, the simulated residual series fit the observed data quite closely.

5.6 Quantile implementation of the JMP and DFL models

The quantile approach nests both the JMP and DFL models.

5.6.1 Quantile implementation of JMP

As before, write the observed distribution of wages at time t as a function of three components: the distribution of worker characteristics, $g_t(X)$, the between-group prices β_t^b for those characteristics, and the within-group prices for those characteristics β_t^b, (suppressing "hats" on the estimated vectors).

The observed change in inequality between any two periods, t and τ, can be decomposed into three components using the following sequential decomposition. Let $\Delta Q_\theta = Q_\theta(f_\tau(w)) - Q_\theta(f_t(w))$ equal the observed change in the θ^{th} wage quantile between periods t and τ. Define

$$\Delta Q_\theta^X = Q_\theta\left(f\left(g_\tau(X), \beta_t^b, \beta_t^w\right)\right) - Q_\theta\left(f\left(g_t(X), \beta_t^b, \beta_t^w\right)\right),$$

as the contribution of changing quantities (labor force composition) to ΔQ_θ. Define

$$\Delta Q_\theta^b = Q_\theta\left(f\left(g_\tau(X), \beta_\tau^b, \beta_t^w\right)\right) - Q_\theta\left(f\left(g_\tau(X), \beta_t^b, \beta_t^w\right)\right)$$

as the marginal contribution of changing between-group prices to ΔQ_θ. And, finally define

$$\Delta Q_\theta^w = Q_\theta\left(f\left(g_\tau(X), \beta_\tau^b, \beta_\tau^w\right)\right) - Q_\theta\left(f\left(g_\tau(X), \beta_\tau^b, \beta_t^w\right)\right)$$

as the marginal contribution of changing within-group prices to ΔQ_θ.

Notice that this decomposition sums to the total observed change: $\Delta Q_\theta^X + \Delta Q_\theta^b + \Delta Q_\theta^w = \Delta Q_\theta$. This is an important advantage over the JMP procedure, in which the 'residual price and quantity component' must be estimated as a remainder term after the other two components are calculated.

5.6.2 Quantile implementation of DFL

Interestingly, the notation above makes it apparent that the DFL procedure is simply the first step of the JMP decomposition above. In particular,

$$\Delta Q_\theta^{DFL} = Q_\theta\left(f\left(g_\tau(X), \beta_t^b, \beta_t^w\right)\right) - Q_\theta\left(f\left(g_t(X), \beta_t^b, \beta_t^w\right)\right) = \Delta Q_\theta^X.$$

Why does DFL only do Step 1? Because the DFL procedure simply reweights the function mapping $X's$ to $w's$ (given by β_t^b, β_t^w in our QR model) using the change in the density of $X's$ between t and τ ($g_t(X)$ to $g_\tau(X)$). Since DFL do not explicitly model 'prices' (β_t^b, β_t^w in the quantile model), they do not take Steps 2 and 3 where these prices are varied.

Similarly, the Lemieux (2004) procedure for reweighting residual densities can be written as

$$\Delta Q_\theta^L = Q_\theta\left(f\left(g_\tau(X), \beta_t^b = 0, \beta_t^w\right)\right) - Q_\theta\left(f\left(g_t(X), \beta_t^b = 0, \beta_t^w\right)\right) = \Delta Q_\theta^X.$$

Unlike the quantile approach, Lemieux estimates β_t^w using OLS. But this difference is unlikely to be important.

5.6.3 Advantages and disadvantages of the quantile decomposition relative to other approaches

An advantage of the QR approach is that it nests JMP, DFL, and all extensions to DFL that have been recently proposed. I would also argue that it handles each of these cases somewhat more transparently than the competing techniques.

A second virtue of QR is that procedure explicitly models the separate roles of quantities, and between- and within-group prices to overall inequality. That is, DFL and extensions never explicitly estimate prices, although these prices are implicit in the tool. By contrast, JMP do estimate prices (both observed and unobserved). But in practice, their residual pricing function does not quite work as advertised unless one conditions the residual distribution ($F_t(\theta|X_{it})$) very finely on all combinations of $X's$.

A third virtue of QR is that it satisfies the adding-up property. That is, if the QR model fits the data well, the sum of the components of the decomposition will add up to the total. (Of course, it is still a sequential decomposition; the order of operations matters.)

Finally, unlike JMP and extensions, QR provides a consistent treatment of between- and within-group prices (there is no 'hybridization' of OLS and logit/probit models).

The QR decomposition has two notable disadvantages.

First, it is parametric. The precision of the simulation will depend on the fit of the QR model which in turn depends on the characteristics of the data and the richness of the QR model. By contrast, the DFL procedure and its extensions never actually parameterize the conditional distribution of wages, $F(w|X)$. Hence, the treatment of $F(w|X)$ in DFL is fully non-parametric. Notably, DFL must parameterize the reweighting function (through the probit/logit). I've never seen any work documenting the amount of slippage induced by that process. It is possible that DFL provides a more precise fit to the distribution than the fully parametric QR model. (It would be nice to test this).

Second, the QR model is computationally intensive. Consider that one must estimate $\sim 100-500$ QR models for each data set, save a large matrix of parameters, and then draw large amounts of data from $g(X)$ and θ to calculate $\hat{w}$ and form counterfactual distributions. Given current computational resources, this is typically burdensome.

6 [Optional] Application of quantile decomposition to U.S. wage structure 1973 - 2003: Lemieux 2006 and Autor, Katz and Kearney 2005

The 2006 *AER* paper by Lemieux on your syllabus proposes a novel explanation for rising residual wage inequality in the U.S. "Using data from the May and Outgoing Rotation Group (ORG) supplements of the CPS, this paper shows that a large fraction of the growth in residual wage inequality between 1973 and 2003 is due to *spurious* [emphasis Autor's] composition effects. These composition effects are linked to the secular increase in the level of experience and education of the workforce, two factors associated with higher within-group wage dispersion. Once these factors are corrected for, I find that residual wage inequality only accounts for a small share of the overall growth in wage inequality. Furthermore, all of the growth in residual wage inequality occurs during the 1980s."

The Autor, Katz, Kearney 2005 paper ("Rising Wage Inequality: The Role of Composition and Prices") is in large part a comment on Lemieux 2006. So, it should be stressed that the important insight that changing labor force composition has contributed to the observed rise of residual inequality is due to Lemieux.

Viewing the same facts, AKK reach quite different conclusions from Lemieux. In particular, they do not concur that "Furthermore, all of the growth in residual wage inequality occurs during the 1980s." Why?

There are two main substantive differences between Lemieux and AKK. First, Lemieux does not (for the most part) distinguish between movements in upper and lower tail residual inequality. Instead, he looks only at the sum of the two (the 90/10 – or the variance of wage residuals, which obviously doesn't distinguish between the upper and lower tail). Why is that problematic? As discussed in class, upper and lower tail inequality only move in tandem from 1979 to 1986. After 1986, they diverge radically, with lower-tail inequality flattening and then compressing and upper-tail inequality rising steadily to the present. Hence, analyzing only the *aggregate* trend in the two may provide an incomplete picture.

In particular, Lemieux finds that the rising education and experience of the labor force can mechanically explain the plateauing of residual inequality during the 1990s. But this inference appears to aggregate over two countervailing forces. The first is the contraction in lower-tail inequality in the 1990s (which as shown above) appears 'due to' compressing residual 'prices' buffered by changing composition. The second is the rise in upper-tail inequality, which appears entirely explained by price changes. Because composition over-explains the former phenomenon and under-explains the latter, it is technically accurate to say that composition can "fully explain" the aggregate trend in residual inequality during the 1990s. But when upper and lower-tail inequality are considered separately – as seems appropriate given their substantial divergence in the last decade – composition does not appear a satisfying explanation for either.

The second main difference with Lemieux (much less important) turns on an issue posed by all sequential decompositions—namely, that the order of operations matters. Lemieux chooses to focus on counterfactual residual inequality trends that hold labor force composition constant at the 1973 level while varying 'prices' over 1973 to 2003. Alternatively, one could have held composition at 1988, 2003, or any other intervening year. As the AKK figures show, this choice matters. Holding composition at the initial 1973 level maximizes the inference that residual inequality fell after 1988. Why? Because the 1973 labor force over-represents high school dropouts and high school graduates and under-represents college grads relative to the labor force of the next three decades. Now, recall that the *compression* of inequality after 1987 is greatest for less educated workers and the *expansion* greatest for more educated workers. Hence, holding composition at the 1973 level maximizes the representation of groups experiencing wage compression and minimizes the representation of groups experiencing wage expansion. As shown in the AKK figures, Lemieux's choice of weights does matter for his conclusions.

[A final point of contrast is that Lemieux's decomposition focuses only on residual inequality. A potential disadvantage of this approach is that the allocation of wage variation into 'between-group' and 'residual' components is inherently arbitrary; it depends upon which $X's$ are included in the conditioning set. By contrast, the sum of between and within-group inequality is invariant to the conditioning set – since it is equal to observed inequality. For this reason, it may be generally more useful to analyze 'between-group' and 'residual' inequality jointly.]

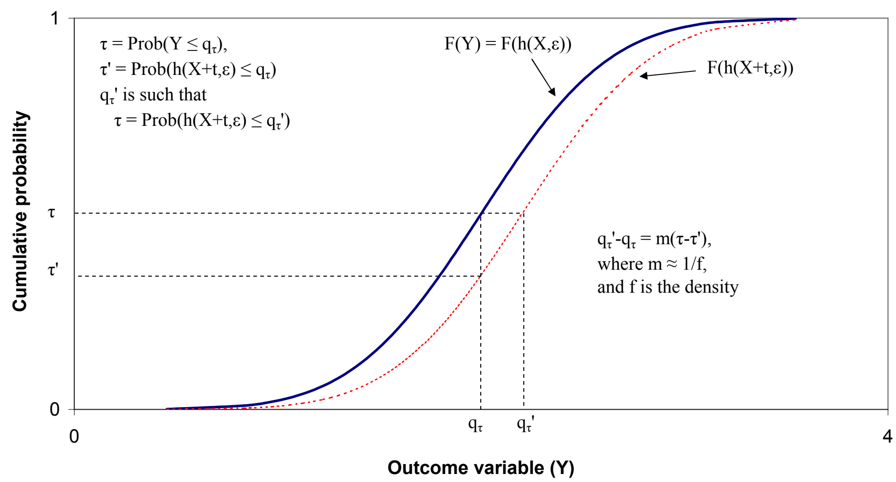
7 Unconditional Quantile Regression

The 2009 *Econometrica* paper, "Unconditional Quantile Regressions" (UCR), by Firpo, Fortin and Lemieux, develops a set of tools for estimating the effect of covariates on the *unconditional* distribution of wages. This is valuable because, as above, we are generally much more interested in the effect of covariates on the unconditional (marginal) distribution of wages than on the conditional distribution. As we saw with Machado-Mata, moving from a conditional quantile regression to an estimate of the effect of covariates on the marginal wage distribution is cumbersome, typically requiring numerical integration. The FFL approach presents a way around this problem.

The discussion below outlines their approach, drawing with permission on (1) lecture notes by Arin Dube of UMass Amherst (Econ 797B: Empirical Methods in Labor Economics, Fall 2011); and (2) recitation notes by Sally Hudson, 14.662 TA in spring of 2013. Sally in turn credits James Gentle of George Mason University for *his* lecture notes on "Sensitivity of Statistical Functions to Perturbations in the Distribution."[3]

7.1 Graphical intuition

Firpo, Fortin, and Lemieux (2009) a method for recovering effects on unconditional quantiles from conditional quantile estimates that is much less computationally intensive than the Machado-Mata approach. Before we dive into the math, let's introduce the intuition graphically.



To fix ideas, suppose we want to estimate the effect of education on the U.S. wage distribution. Imagine that the blue line in the above graph plots the CDF of log wages, Y. That means the vertical axis measures quantile levels $\tau \in [0,1]$ and the horizontal axis measures the corresponding wage values at each quantile. Let's say the hash marked τ is the median so that q_τ is the median wage, which is currently about \$27,000.

Now suppose we perturbed the distribution of educational attainment in the population – perhaps by giving everyone an additional year of schooling. What would we expect to happen to the wage distribution? In partial equilibrium, earnings should rise throughout the wage distribution. That means that \$27,000 would now correspond to a lower quantile in the wage distribution, labeled τ' above. Indeed, at every wage value we would expect to find fewer people with wages below that value. This shift is represented by the red CDF,

[3]Link to Gentle's Notes

which lies everywhere below the blue CDF. By the same token, for any fixed quantile (like the median τ) the new wage value ($q_{\tau'}$) will be larger than the old wage value (q_τ).

The goal of unconditional quantile regression is to estimate quantities like $q'_\tau - q_\tau$, the change in median wages associated with the change in education. FFL's insight is that if the blue CDF and red CDF have roughly the same slope around q_τ, then

$$(\text{slope of blue CDF at } q_\tau) \times (q'_\tau - q_\tau) = \tau - \tau'$$

And what is the slope of the CDF? It's the probability density function f_Y, which is something we know how to estimate! We can therefore back out the effect on the median as follows.

$$q'_\tau - q_\tau = \frac{\tau - \tau'}{f_Y(q_\tau)}$$

This is the intuition behind the unconditional quantile regression method that follows.

7.2 Influence functions

Now for a quick mathematical digression. To understand FFL's procedure, we need to discuss a class of functions called influence functions. Influence functions measure the sensitivity of a distributional statistic, like the mean or median, to a small change in the data's distribution. Taking the derivative of one function with respect to another function is a broad class of problems in analysis, but the good news is that cumulative density functions are a rather narrow class of functions. They're bounded on the unit interval, and they're monotonically increasing. These restrictions allow for a nice characterization of the functional derivative known as the Gateaux derivative.

7.2.1 The Gateaux derivative

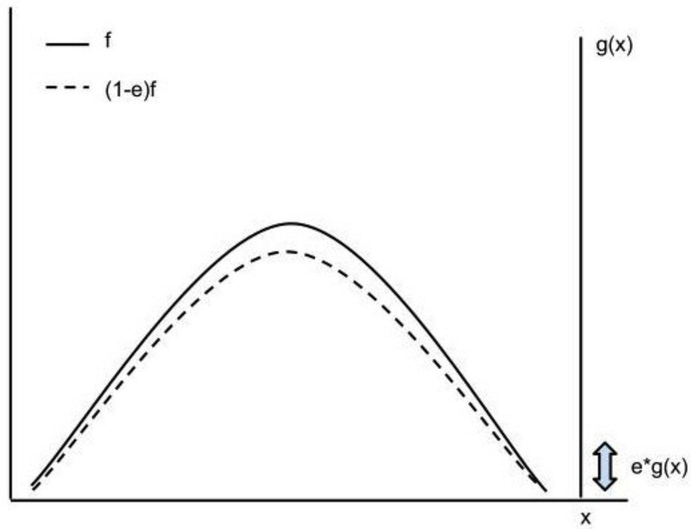
Let T be a statistic of a distribution F and let G be an alternate distribution. The Gateaux derivative captures the change in the value of T as we perturb F to look more like G. Formally, the Gateaux derivative is written

$$L(G; T, F) \quad = \quad \lim_{\epsilon \to 0} \frac{T\left[(1-\epsilon)F + \epsilon G\right] - T(F)}{\epsilon}$$

Note that the first term in square brackets is just a convex combination of F and G. As $\epsilon \to 0$, the expression puts more weight on F and less weight on G.

Influence functions are Gateaux derivatives in which the distribution G is just a point mass at some value x. That is,

$$G(\mu) = \begin{cases} 0 & \text{for } \mu < x \\ 1 & \text{for } \mu \geq x \end{cases}$$



The picture above illustrates the idea of perturbing a distribution by a point mass. The solid curve plots the PDF of the reference distribution F. Imagine that we shaved an ϵ mass off the top of f and piled it up at the point x on the right. As $\epsilon \to 1$, the PDF of f would flatten and disappear and the mass at x would approach a point mass. Now consider the same change in reverse. Imagine we started with a point mass at x and took $1-\epsilon$ of that mass and spread it out over the distribution of f. As $\epsilon \to 0$, the point mass would shrink and the dashed curve would converge to f. This is conceptually what's happening in the definition of the influence function. If we let δ_x denote a point mass at x, the influence function for statistic T and distribution F is

$$IF(x; T, F) \quad = \quad \lim_{\epsilon \to 0} \frac{T\left[(1-\epsilon)F + \epsilon\delta_x\right] - T(F)}{\epsilon}$$

In essence, the influence function asks, "What would happen to my statistic T if I added a single observation at x to the distribution F?" Recall that a single point has no mass in a continuous distribution, which is why we obtain this result by letting $\epsilon \to 0$.

7.2.2 Example: The Influence function of the mean

Here's a quick example of an influence function derivation. Let $T(F)$ be the mean of F. The influence function is then

$$
\begin{aligned}
IF(x;T,F) &= \lim_{\epsilon\to 0}\frac{\mathbb{E}\left[(1-\epsilon)F+\epsilon\delta_x\right]-\mathbb{E}\left[F\right]}{\epsilon}\\
&= \lim_{\epsilon\to 0}\frac{(1-\epsilon)\mathbb{E}\left[F\right]+\epsilon\mathbb{E}\left[\delta_x\right]-\mathbb{E}\left[F\right]}{\epsilon}\\
&= \lim_{\epsilon\to 0}\frac{-\epsilon\mathbb{E}\left[F\right]+\epsilon x}{\epsilon}\\
&= x-\mu_F
\end{aligned}
$$

In finite samples, the formula below shows how the influence function can be used to determined the effect of a new data point on an estimated statistic.

$$
T_n \approx T_{n-1}+\frac{1}{n}IF(x;T,F_{n-1})
$$

To see this in action, suppose our initial sample is just $\{3,4,5\}$. The mean of this sample is 4. Now suppose we add an observation at 20. Then the new mean is

$$
\begin{aligned}
\hat{\mu}_{n-1}+\frac{1}{n}\left(x-\hat{\mu}_{n-1}\right) &= 4+\frac{20-4}{4}\\
&= 8
\end{aligned}
$$

which you can verify is the mean of the sample $\{3,4,5,20\}$.

7.2.3 The influence function of the quantile

The influence function of the quantile is only slightly more complicated.[4]

$$
\mathrm{IF}(Y_i;q_\tau,F_Y)=\frac{\tau-\mathbf{1}\left\{Y_i\leq q_\tau\right\}}{f_Y(q_\tau)}
$$

The intuition is similar to the graphical argument made in Section 7.1. For a given τ, if we observe a new data point below q_τ we adjust our estimate of the quantile downward. We then scale that adjustment by the slope of the CDF, which translates a change in $\tau's$ to a change in Y.

[4] I have temporarily removed the derivation of this influence function from these notes as it may be used on the upcoming problem set.

7.2.4 The Recentered Influence Function (RIF)

The key to FFL's unconditional quantile estimation method is a clever transformation of the influence function. They define the recentered influence function (RIF) as the sum of the influence function and the original statistic.;

$$\begin{aligned}\mathrm{RIF}(Y_i; q_\tau, F_Y) &= q_\tau + \mathrm{IF}\left(Y_i; q_\tau, F_Y\right) \\ &= q_\tau + \frac{\tau - \mathbf{1}\left\{Y_i \leq q_\tau\right\}}{f_Y(q_\tau)}\end{aligned}$$

Note that the expectation of the RIF for any quantile q_τ is just the quantile itself.

$$\begin{aligned}\mathbb{E}\left[\mathrm{RIF}(Y_i; q_\tau, F_Y)\right] &= q_\tau + \frac{\tau - \mathbb{E}\left[\mathbf{1}\left\{Y_i \leq q_\tau\right\}\right]}{f_Y(q_\tau)} \\ &= q_\tau + \frac{\tau - \tau}{f_Y(q_\tau)} \\ &= q_\tau \qquad (25)\end{aligned}$$

This observation is going to help us to overcome the linearity problem discussed above.

7.3 RIF Regression

We now have the tools we need need to estimate the response of an unconditional quantile to a change in explanatory variables. As in both OLS and CQR, we start by estimating a linear approximation to a conditional function, in this case the conditional RIF.

$$\mathbb{E}\left[RIF\left(Y_i; q_\tau, F_Y\right) \mid X_i\right] = X_i'\beta_\tau + \epsilon_i \qquad (26)$$

We can then use the identity in (25) to show that these coefficients will also capture effects on the unconditional quantile function, which is what we're really after.

$$\begin{aligned}q_\tau &= \mathbb{E}\left[\mathrm{RIF}(Y_i; q_\tau, F_Y)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[RIF\left(Y,; q_\tau, F_Y\right) \mid X_i\right]\right] \\ &= X_i'\beta_\tau\end{aligned}$$

To estimate β_τ, we start by rearranging the conditional RIF in terms of things we observe in the data.

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{RIF}(Y_i;q_\tau,F_Y)\mid X_i\right] &= q_\tau + \frac{\tau - \mathbb{E}\left[\mathbf{1}\left\{Y_i \le q_\tau\right\}\mid X_i\right]}{f_Y(q_\tau)} \\
&= q_\tau + \frac{\tau - \left(1 - \Pr\left\{Y_i > q_\tau \mid X_i\right\}\right)}{f_Y(q_\tau)} \\
&= \left(q_\tau + \frac{\tau - 1}{f_Y(q_\tau)}\right) + \frac{1}{f_Y(q_\tau)} \cdot \Pr\left\{Y_i > q_\tau \mid X_i\right\} \\
&= c_{0\tau} + \frac{1}{f_Y(q_\tau)} \cdot \Pr\left\{Y_i > q_\tau \mid X_i\right\} \qquad (27)
\end{aligned}
$$

Substituting (26) into (27) gives

$$
\begin{aligned}
c_{0\tau} + \frac{1}{f_Y(q_\tau)} \cdot \Pr\left\{Y_i > q_\tau \mid X_i\right\} &= X_i\beta_\tau + \epsilon_i \\
\Pr\left\{Y_i > q_\tau \mid X_i\right\} &= -c_{o\tau} + X_i\beta_\tau f_Y\left(q_\tau\right) + \epsilon_i \qquad (28)
\end{aligned}
$$

And equation (28) is something we can estimate! Here's how:

1. Generate a dummy variable D_i that indicates whether person i's wage exceeds the chosen quantile. That is,

$$
D_i(Y_i) = \begin{cases} 0 & \text{for } Y_i \le q_\tau \\ 1 & \text{for } Y_i > q_\tau \end{cases}
$$

   This is the left hand side variable in (28).

2. Run an OLS regression of D_i on a constant and the vector of covariates X_i.

$$
D_i = \gamma_0 + X_i\gamma_1 + \nu_i
$$

   Note that $\gamma_1 = \beta_\tau f_Y(q_\tau)$ in (28), which is just the marginal effect of X on the fraction of outcomes above the threshold q_τ.

3. Generate a kernel density estimate of f_Y to obtain $\hat{f}_Y(q_\tau)$.

4. Divide $\hat{\gamma}_1$ by $\hat{f}_Y(q_\tau)$ to obtain $\hat{\beta}_\tau$.

$$
\hat{\beta}_\tau = \frac{\hat{\gamma}_1}{\hat{f}_Y(q_\tau)}
$$

That's it! This is the estimated effect of a change in the covariates on the unconditional quantile of Y.

Note that more generally for any cutoff y^c, we can create a dummy variable $\tilde{y}_c = 1(y > y^c)$. We can then regress $\tilde{y}_c$ on $X'\beta_\tau$ and get $\hat{\beta}_c$. This regression estimates a well defined quantity, the marginal effect of X on the fraction of outcomes above a cutoff. Inverting that relationship is the key to both FFL's approach and to Chernozhukov et al. (2009), who propose globally inverting distributions rather than using FFL's localized approach.

7.4 Limitations of RIF Regression

RIF regression methods are only as good as the kernel density estimate of f_Y. For variables that are smooth by construction, like standardized test scores, this is not such a problem, but for variables where there is considerable heaping, like wages, the estimates of f_Y may depend a lot on subjective choices about the smoothing factor. Note that when values of $\hat{f}_Y(q_\tau)$ are close to zero, the difference between, say, $\hat{f}_Y(q_\tau) = 0.10$ and $\hat{f}_Y(q_\tau) = 0.05$ leads to a doubling of the estimated treatment effect, so discrepancies that seem small in an absolute sense can translate into big differences in outcomes.

RIF regression also relies on the assumption that we can locally invert f_Y in some neighborhood of q_τ. Chernozhukov et al. (2009) show that we can sometimes inverted the density function globally rather than locally, which eliminates the need for this assumption. Global inversion is very computationally intensive, however, and sometimes intractable.

And, as with all of the decomposition methods we've discussed, RIF regressions produce partial equilibrium results. They don't allow for prices to respond to changes in the distribution of X's, which we know is probably unrealistic.

8 The Contribution of Tasks, Unions, and Labor Market Composition to Wage Structure Changes

The 2011 working paper by FFL called "Occupational Tasks and Changes in the Wage Structure" applies the FFL unconditional quantile model to examine the role that three sets of factors have played in the evolution of the U.S. *male* wage structure (no idea why they limit to males) over the last several decades (though their focus is on 1989–2001). These factors are changes in labor market composition (education, experience), changes in unionization, and changes in the returns to various tasks. All three of these factors are allowed to affect the wage structure through both a compositional (quantity) effect and a wage structure (price) effect. The quantity effect corresponds to the observations that some $X's$ may be associated with higher wage dispersion (or more complex quantile effects), for example, there is higher wage variance among experienced college-educated workers than among experienced high-school educated workers. The wage structure (price) effect refers to the change in the returns to those same characteristics (e.g., education, experience), which may also affect wages at any and all quantiles. The distinction between composition and price effects is relevant because a change in the distribution of $X\prime s$ can affect the shape of the wage distribution without any change in the returns to those $X's$, and more conventionally, a change in the returns to a set of $X\prime s$ can alter the shape of the wage distribution without any change in the distribution of $X's$.

Other than their use of the UCR, the main conceptual object introduced in FFL 2011 is what they refer to as a Roy model of occupational task pricing. Specifically, FFL propose that the wage paid to individual i in occupation j at time t can be written as:

$$w_{ijt} = \theta_{jt} + \sum_{k=1}^{K} r_{jkt} \times S_{ik} + u_{ijt},$$

where r_{jkt} is the return to task k in occupation j in year t and s_{ik} is the skill of worker i in task k. Note that there is no 'law of one price' for tasks in this model; tasks have different values in different occupations. Hence, as in a Roy model, workers will *not* be indifferent across occupations since some occupations will offer strictly higher wages than others for the skill bundle that they possess.

This naturally leads to an array of questions on how equilibrium task prices are set, how workers self-select across occupations, what equilibrium conditions govern the system, etc. Without very strong additional assumptions, it will clearly be quite difficult to get any firm predictions about equilibrium pricing and self-selection in this model, however. FFL do not attempt to tackle these problems. Rather, they implicitly take as exogenous the pricing of tasks and, additionally, take as fixed (or at least ignorable) the self-selection of workers into occupations with differing task demands and prices. This latter issue (self-selection) seems potentially first order since if task prices are changing, as FFL posit, this should cause endogenous reallocation of workers across occupations in response to changing comparative advantage.

FFL's model is probably best understood as a statistical statement rather than an economic model since both partial equilibrium considerations (self-selection of workers across occupations according to comparative advantage) and general equilibrium conditions (simultaneous determination of worker assignments and task prices) are not actually considered. While this statistical model may be suitable for FFL's paper, it's probably a bit of a stretch to call it a Roy model since it ignores the economic phenomenon—endogenous assignment of workers to jobs—that Roy's model was designed to interpret.

The empirical leverage that the FFL statistical model offers is as follows. Rewriting the equation above in first-differences:

$$\Delta w_{ij} = \Delta\theta_j + \sum_{k=1}^{K} \Delta r_{jk} \times S_{ik} + \Delta u_{ij}.$$

Using a simple linear approximation, one can express the relationship between the change in the wage and its initial level as:

$$\Delta w_{ij} = \tilde{a}_j + \tilde{b}_j w_{ij0} + e_{ij},$$

if one assumes (restrictively) that the skill components S_{ik} are uncorrelated, it's a straightforward to calculate that:

$$\tilde{b}_j = \frac{\text{Cov}\left(\Delta w_{ij}, w_{ijo}\right)}{\text{Var}\left(w_{ijo}\right)} = \frac{\sum_{k=1}^{K} r_{jk0}\Delta r_{jk}\sigma_{jk}^2}{\sum_{k=1}^{K} r_{jk0}^2\sigma_{jk}^2 + \sigma_{ujo}^2}.$$

Thus, the 'slope' relating wage changes to wage level will tend to steepen in occupation j if task prices r_{jk} in that occupation rise. This basic idea is developed more fully in the Technical Appendix of FFL 2011.

Empirically, FFL estimate their models in two stages. A first is to estimate models for quantiles of wage changes in each occupation:

$$\Delta w_j^q = \tilde{\alpha}_j + \tilde{\beta}_j w_{j0}^q + \lambda^q + \varepsilon_j^q,$$

where the $q's$ refer to percentiles of the wage distribution. The $\tilde{\alpha}$'s measure 'between' occupation wage changes while the $\tilde{\beta}$'s measure within occupation changes in variance.

FFL then regress the estimated $\tilde{\alpha}_j$'s and $\tilde{\beta}_j$'s on occupational tasks, which they interpret using the following approximations:

$$\tilde{\beta}_j \approx \frac{\sum_{k=1}^{K}\left(r_{jk0}\Delta r_{jk}\right)\cdot\sigma_{jk}^2}{\sigma_{j0}^2} + \frac{\Delta\sigma_{jg}^2}{2\sigma_{j0}^2},$$

and

$$\tilde{\alpha}_j = \Delta\delta_j + \sum_{k=1}^{K} \Delta r_{jk}\bar{S}_{jk}.$$

Thus, loosely, this model says that the intercepts and slopes of occupations with rising task prices will rise, and vice versa for occupations with falling task prices. Under the (unstated) hypothesis that although task prices differ across occupations, their changes move similarly across occupations (that is, the price of routine tasks declines in all occupations simultaneously), this model makes predictions for the evolution of slopes and intercepts of occupations as a function of their initial routine, abstract and manual task content (though these are not the terms that FFL employ).

9 Conclusion

Ultimately, the DFL, JMP, and MM wage density decomposition approaches to analyzing counterfactuals can only be slightly convincing at best given that they apply partial equilibrium tools to analysis of general equilibrium problems. Nevertheless, these techniques are widely used, and it is worth understanding them.