Measurement and Testing of Inequality from Time Series of Deciles with an Application to U.S. Wages

Andrew Harvey and Jared Bernstein
Faculty of Economics and Politics, University of Cambridge, U.K.
and Economic Policy Institute, Washington, D.C., U.S.A.

January 27, 2000

Abstract

This article uses unobserved components time series models to capture the underlying trends in the quarterly deciles of US hourly wages. Tests of stability and divergence are suggested as a means of assessing changes in inequality. The decrease in the wage gender gap is examined and the impact of changes in the minimum wage is assessed.

KEYWORDS: Co-integration, Common trends, Kalman filter smoother, Minimum wage, Percentiles, Stochastic trend, Unobserved components.

JEL classification: C22, D63

1. Introduction

The growth of economic inequality has been a major concern of economists and policy makers. In the US a substantial literature has emerged examining the increased dispersion in the distributions of wages, incomes, and wealth; see Katz and Autor (1999) for a recent review of wage inequality and Mishel *et al*, (1999, Chapters 1 and 5) for a review of recent work on income and wealth inequality. Since the largest component of income for most (non-elderly) families derives from

their earnings, much analysis has focused on the increase in earnings inequality.[1] To distinguish the impact of changes in the variability of hours worked from those of hourly wages, labor economists have often focused on the dispersion of the latter. This approach has theoretical justification as well, since the microeconomic theory of wage determination is usually cast in terms of the hourly wage. In US labor markets, virtually all of the evidence has been presented in terms of annual wage trends. A typical approach uses annual data from the Current Population Survey (CPS) to calculate various indicators of wage inequality. For example, Katz and Autor (1999) use data from the March CPS to calculate the annual growth in the log difference between the 90th and 10th percentile wages. While this approach has been informative, the relatively small number of observations for annual percentile time series has apparently limited the scope for applying statistical modelling techniques to extract underlying trends and to examine the relationship between them. For the same reason, the literature is quiet on what represents a statistically meaningful increase in measures of inequality.

This study uses a newly constructed quarterly data set on hourly wage deciles. However, while yielding more information, quarterly data also introduce some new measurement challenges. Seasonal effects may be present and the observations will tend to be more 'noisy' because irregularities and survey errors have not been averaged out over the year. What is needed is a method which efficiently extracts the underlying trend from such components. Structural time series models are set up precisely to accomplish this task. Univariate models provide an excellent means of displaying the stylized facts surrounding movements in wage inequality, while multivariate models suggest a framework for a more rigorous analysis based on recent ideas on nonstationarity, common factors and co-integration. Although this theory is now well established, it has not, to our knowledge, been applied to percentile time series, nor have the tests we propose been used in this context. Our experience with these methods may therefore be useful to other researchers working with time series data on the distributions of economic variables.

The structure of the paper is as follows. The data are described in section 2. This is followed, in section 3, by a description of how univariate structural (unobserved component) time series models can be used to extract underlying trends from seasonal series. Section 4 then fits the model to the data on deciles of male wages. Section 5 looks at how inequality can be captured by various

---

[1]According to the Committee on Ways and Means, 1992, earnings comprised 76% of total family income from all sources for families in the middle quintile of the income distribution in 1989.

'contrasts' such as the difference between, or ratio of, two deciles. Tests of stability and divergence (or convergence) are proposed and applied to the data. These tests may yield information which would not be apparent from simply 'eyeballing' the unadjusted data. Multivariate models are introduced in section 6 and the link with common factors and co-integration is explored in some detail. A general statistical test for wage dispersion, based on eight contrasts, is given and the results of fitting multivariate models to wage deciles are reported.

Section 7 explores the changing relationship between male and female wages, while section 8 goes on to look at the impact of minimum wage legislation. The conclusions are presented in section 9.

2. The Data

The quarterly hourly wage data are derived from the Current Population Survey, a monthly survey conducted by the US Bureau of the Census. Each month, approximately 50,000 households are interviewed regarding their labor market status in the middle week of the month. Since 1979, one-quarter of the sample, called the outgoing rotation groups (ORG), are asked questions about their earnings that week.[2] The survey also collects demographic information on these workers and provides a weighting factor to make the ORG sample nationally representative. Our sample includes only 18 to 64 year olds, in order to exclude those whose youth or age might preclude steady labor force attachment. The earnings questions are asked of wage and salaried workers (excluding the self-employed) in both the private and pubic sectors. Respondents are also asked whether they are paid by the hour in some other format.[3] For hourly workers, we simply take the hourly wage; for others, we divide their weekly earnings by their usual weekly hours. More detail on the construction of the wage variable can be found in the data appendix of Mishel *et al* (1999). A fairly extensive revision of these earnings variables was

---

[2] The CPS survey is administered as a "rolling panel." Households entering the survey are interviewed for four consecutive months, leave the survey for eight months and are then back in for four months (the same calendar months as the first four). The earnings questions are asked only to those households either in their fourth or eighth month in the survey. Since these households are leaving the rotation, either temporarily (for those in the fourth month in survey) or permanently (for those in their eighth survey month), the earnings sample is referred to as the outgoing rotation group.

[3] Prior to 1994, respondents had to either report an hourly wage or weekly earnings. From 1994 forward, pay periods other than weekly were introduced. See Polivka (1998) and Bernstein and Mishel (1997) for a discussion of the impact of these changes on measured wage trends.

introduced in 1994. Bureau of Labor Statistics (BLS) statisticians, in particular Polivka (1998), provide detailed analysis of the impact of the change in measured earnings. This work, along with the analysis in Bernstein and Mishel (1997), suggests that the impact of the revisions does not create a significant inconsistency in the series. The CPS earnings data are top-coded to preserve confidentiality, and in other work with these data, such as Mishel *et al* (1999) various adjustments are made for this problem. However, the current analysis examines hourly wage values up to, but not above, the 90th percentile of the gender-specific wage distribution. Since topcoded cases are well above this cutoff (typically above the 97th percentile), this does not affect our analysis. In order to control for obvious outliers, we exclude hourly wages below $0.50 and above $100 in 1989 dollars.

The data are deflated using the CPI-U-X1, an experimental BLS deflator which corrects the overstating of price growth in the more commonly used CPI-U series in the late 1970s and early 1980s (the US Census Bureau also uses the X1 in deflating their income series). Although the X1 series is also susceptible to recent critiques of the CPI, our interest here is in the gap between wages at different deciles, and the magnitude of the gap is unaffected by the deflator. As noted in West (undated) and Polivka (1998), wage quantiles derived from survey data tend to "clump" around commonly reported values, such as (in the case of hourly wages), $0.50 and $1.00 intervals. Over time, these spikes in the wage distribution tend to persist for a number of periods before jumping to the next spike. Thus, a particular wage quantile computed from the raw data might be $7.00 for a few consecutive quarters before eventually jumping to $7.50. Analysts of movements in wage quantiles need a method to control for this aspect of reporting bias. We use an interpolation method recommended by West and used by the US BLS in their work; see Mishel *et al* (1999, Appendix B) for a detailed explanation.

Figure 1 shows some of these interpolated hourly wage deciles, in logged real 1998 dollars, for all workers, from 1979q1 to 1999q3. In order to avoid clutter, the figure shows the 10th, 30th, 50th, 70th, and 90th percentiles; logarithms are commonly used in this type of analysis since they make the gaps more equal and focus attention on relative differences. While the figure combines male and female wages, we follow the more common practice of examining wage trends by gender in much of what follows. The gender-specific trends are influenced by changes in labor supply that were much greater for women over this period, as well as occupational segregation, which actually diminished slightly. Putting aside for a moment the question of growing inequality, which is somewhat obscured by the magnitude of the scale, the trends in figure 1 appear flat or falling slightly

for the middle three series, falling more steeply for the first decile and drifting upwards for the ninth decile. Starting around 1996, tight labor markets in the US began to generate fairly broad based wage growth. The series have numerous spikes and dips, but the structural time series models enable us to filter out seasonal and irregular effects, to test whether apparent movements in inequality are statistically significant and to pinpoint the impact of changes in the minimum wage.

3. Univariate models for trend extraction

This section sets out the statistical models used in the analysis and reviews how to test against the presence of nonstationary unobserved components.

3.1. Stochastic trends

The *local level* model is

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \quad t = 1, ..., T \tag{3.1}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim NID(0, \sigma_\eta^2), \tag{3.2}$$

where the irregular and level disturbances, ε_t and η_t respectively, are mutually independent and the notation $NID\left(0, \sigma^2\right)$ denotes normally and independently distributed with mean zero and variance σ^2. When σ_η^2 is zero, the level is constant.

The *local linear trend* model is more general in that the trend component, μ_t, has a stochastic slope, β_t. Thus

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2), \end{aligned} \tag{3.3}$$

where the irregular, level and slope disturbances, ε_t, η_t and ζ_t respectively, are mutually independent. When $\sigma_\zeta^2 = 0$ the slope is fixed and the trend reduces to a random walk with drift

$$\mu_t = \mu_{t-1} + \beta + \eta_t. \tag{3.4}$$

Setting σ_η^2 to zero gives an integrated random walk trend, which when estimated tends to be relatively smooth. In fact it is equivalent to a cubic spline; see

5

Wecker and Ansley (1983). The model is often referred to as the 'smooth trend' model. The signal-noise ratio, $q_\zeta = \sigma_\zeta^2 / \sigma_\varepsilon^2$, determines the degree of smoothness.

Trend components are extracted by the Kalman filter and smoother (KFS). This yields the optimal estimator of the trend in all time periods, together with its RMSE. The estimator of the trend can be regarded as a weighted average of adjacent observations, that is

$$\widetilde{\mu}_{t|T} = \sum_j w_{j,t} y_{t+j}, \quad t = 1, ..., T.$$

The weights, $w_{j,t}$, are asymmetric near the beginning and end of the series, but near the centre of a large sample they are symmetric and independent of t. Harvey and Koopman (1999) examine the weights assigned to observations to extract various trends and compare them with the kernels typically used in nonparametric trend extraction. Figure 2 shows the implicit symmetric kernels, that is the time invariant weights, for the smooth trend model with four different signal-noise ratios.

Prior to extracting trends, maximum likelihood (ML) estimates of the parameters, that is variances such as σ_ζ^2 and σ_ε^2, are computed by putting the model in state space form (SSF) and applying the Kalman filter; see Harvey (1989). all the calculations reported here were carried out using the STAMP package of Koopman *et al* (1995).

3.2. Serial correlation

Incorporating a serially correlated component, such as a stochastic cycle, into the model can yield a trend which can be more appealing insofar as it changes more slowly. Harvey and Jaeger (1993) fit such a model to real US GNP in order to extract the long-run trend. In the present context a cycle may not correspond to the business cycle, but may pick up other variations. A simple first-order autoregressive, AR(1), model may also be useful in this respect.

Serial correlation may also be induced by the rotating sample design. In a given quarter, a little less than half of those asked about earnings will have been asked in the same quarter of the previous year (the proportion will be less than one-half because of attrition). This will induce correlation at lag four, which could perhaps be modelled by a fourth-order moving average, MA(4), with gaps at the non-seasonal lags. However, unlike the measurement of employment status, where the implied error process can be deduced, as in Pfeffermann (1991), it seems

difficult to determine the moving-average coefficient. While an MA(4) could be added to the model and the parameter estimated along with the others, it seems unlikely that omitting it will distort the trend in any significant way and we found no evidence of significant fourth-order autocorrelation in the residuals. We have not, therefore, attempted to incorporate this feature in our model.

3.3. Stationarity tests

In the Gaussian random walk plus noise model, (3.1) and (3.2), the locally best invariant (LBI) test of the null hypothesis that $\sigma_\eta^2 = 0$, against the alternative that $\sigma_\eta^2 > 0$, can be formulated as

$$\eta = T^{-2} \sum_{i=1}^{T} \left[\sum_{t=1}^{i} e_t \right]^2 / s^2 > c, \tag{3.5}$$

where $e_t = y_t - \overline{y}$, $s^2 = T^{-1} \sum_{t=1}^{T} (y_t - \overline{y})^2$ and c is a critical value; see Nyblom and Mäkeläinen (1983). The test can also be interpreted as a one-sided Lagrange multiplier (LM) test. The asymptotic distribution of the statistic is Cramér-von Mises, for which the 5% critical value is 0.461.

If the trend is a random walk plus drift, as in (3.4), it becomes deterministic when $\sigma_\eta^2 = 0$. Thus

$$y_t = \mu_0 + \beta t + \varepsilon_t, \qquad t = 1, ..., T. \tag{3.6}$$

The test statistic, η_2, is as in (3.5) except that it is formed from the OLS residuals from a regression on a constant and time. The asymptotic distribution is a second level Cramér-von Mises distribution for which the 5% critical value is 0.149.

If ε_t is any indeterministic stationary process, rather than white noise, the asymptotic distribution of the test statistic under the null hypothesis remains the same if s^2 is replaced by a consistent estimator of the long-run variance. Kwiatkowski *et al* (1992) construct such an estimator nonparametrically. Serial correlation can also be handled parametrically by using the parameters obtained from an unrestricted model to construct a test statistic from the innovations or 'smoothing errors' obtained from running the Kalman filter with σ_η^2 set to zero; see Harvey and Streibel (1997).

A different test is obtained when the alternative is a smooth trend and the null is $\sigma_\zeta^2 = 0$. Nyblom and Harvey (2000a) derive this test but then go on to show that it has no more power than η_2.

Note that unit root tests, such as augmented Dickey-Fuller are inappropriate here as the null of stationarity is the hypothesis to be tested. In any case an autoregressive approximation may not be very satisfactory.

3.4. Seasonality

The basic structural model consists of trend, seasonal and irregular components, that is

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad t = 1, ..., T. \tag{3.7}$$

The trigonometric form of stochastic seasonality has proved effective in modelling a wide variety of seasonal movements; see Harvey (1989, ch 2). The specification is

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}, \tag{3.8}$$

where s is the number of seasons and each $\gamma_{j,t}$ is generated by

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \qquad \begin{matrix} j = 1, \ldots, [s/2], \\ \\ t = 1, \ldots, T, \end{matrix} \tag{3.9}$$

where $\lambda_j = 2\pi j/s$ is frequency, in radians, and $\omega_{j,t}$ and $\omega_{j,t}^*$ are two mutually uncorrelated white noise disturbances with zero means and common variance σ_ω^2, which is the same for all j. For s even $[s/2] = s/2$, while for s odd, $[s/2] = (s-1)/2$. For s even, the component at $j = s/2$ collapses to

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \omega_{j,t}, \qquad j = s/2. \tag{3.10}$$

If the seasonal pattern is of interest in itself it may be extracted by the KFS. It evolves over time, unless $\sigma_\omega^2 = 0$ in which case it is deterministic. Canova and Hansen (1995) show how a test of this hypothesis may be carried out along similar lines to the tests against stochastic trends. However, whether the seasonal pattern evolves over time is a secondary issue in the present context. [4] The essential point is that it can handled within the overall model so that attention can be focused on the trend.

---

[4]It could be argued that some of the seasonality comes from the deflator. However, the deflator we used, the CPI-U-X1, appears to have no significant seasonal component.

4. Decile trends

We now use the techniques of the previous section to present the stylized facts of the movements in the wage distribution. At this stage we confine ourselves to males, firstly because the lower deciles appear to be little affected by minimum wage legislation and secondly because the female distribution partly reflects increased female participation in the labor force.

Let $D_j(j)$ and $d_t(j)$, $j = 1, ..., 9$ denote the $j-th$ decile and its logarithm at time $t = 1, ..., T$. Using logarithms focuses attention on relative, rather than absolute, inequality. However over the period covered by our data there are no big increases in the real wage and so there is little difference in fit between levels and logarithms. In any case having fitted a model in logarithms we can always take antilogarithms of an extracted component. The implications for testing are discussed in sub-section 5.3.

Fitting the BSM, that is (3.7) with the unrestricted trend as in (3.3), to the median, $d_t(5)$, gave the summary statistics shown in table 1. The standard deviation of the one-step ahead prediction errors is denoted 's.e.' while $\log L$ is the maximized log-likelihood. The diagnostics are $r(1)$, first-order residual autocorrelation, and $Q(P, f)$, the Box-Ljung statistic based on the first P residual autocorrelations and assumed to have a χ_f^2 distribution in a correctly specified model. Further details are in the STAMP manual. There is no indication of any misspecification. The only problem is that the extracted trend is not particularly smooth. When the trend is restricted to be a random walk plus drift (RWD), the fit is almost as good, as indeed it is if the drift is omitted (RW), but again the trend is not smooth.[5]

Setting the level variance, σ_η^2, to zero gives an integrated random walk trend (IRW). The fit is slightly worse, with $r(1)$ large, though the Box-Ljung statistic is still satisfactory. Theoretically the hypothesis that σ_η^2 is zero could be rejected as the LR statistic is 4.30 while the 5% critical value is 2.71 for a parameter on the boundary of the parameter space; see Harvey (1989, p 248-9). The compensating advantage is that the trend is smooth and gives a clear indication of the movements in the series. Overall this is our preferred model for the study at hand. In what follows it will be referred to as the *smooth trend BSM*. Figure 3 shows the trend, together with the slope (the quarterly growth rate), the irregular and the seasonal.

---

[5] Note that $\log L$ for RW is not directly comparable with the other $\log L's$ as the state vector has one nonstationary element fewer.

The seasonal variance, σ_ω^2, is zero[6] indicating that the seasonal pattern is constant. The slight first-order serial correlation can be removed by replacing the irregular by an $AR(1)$. This gives a likelihood almost exactly equal to that of the BSM but it has one more parameter. The AR coefficient is only 0.3 and the difference in the trend is difficult to detect by eye.

Other deciles give similar results, sometimes with a relatively better fit for the smooth trend specification. An AR(1) or cycle can make a difference, but overall our preference is to fit the BSM and constrain the trend to be smooth. Table 2 shows the results of fitting smooth trend BSMs to the odd deciles 1,3,5,7,9. The parameter estimates, goodness of fit statistics and diagnostics are close to each other with the possible exception of the $7 - th$ decile.

The seasonal variance was estimated as zero for all deciles. Although the seasonal patterns are relatively small they are highly significant. The smallest value of the joint test statistic, asymptotically distributed as χ_3^2 under the null of no seasonality, is 16.03, for which the probability value is 0.0011. Table 3 shows the percentage deviations from underlying level. These are never more than 1%. The pattern for the first decile is slightly different to the others but not by a great deal.

5. Changing inequality

The contrasts between different deciles point to various aspects of inequality. We will define a contrast as a linear combination of deciles, that is

$$c_t = \sum_{j=1}^{9} \alpha_j d_t(j) = \boldsymbol{\alpha}' \mathbf{d}_t \tag{5.1}$$

Changing inequality may be captured by modeling contrasts and extracting the trend. The model specification is the same as for deciles. Indeed if two deciles can each be modelled by a BSM, the contrast is a BSM and the parameters can be deduced from the individual parameters and the covariances between the disturbances in the two series. Section 6 shows how this result generalises to more than two series and to several contrasts.

---

[6]Initial estimation actually gives it a small positive value, but re-estimating with it fixed increases the likelihood.

The simplest contrasts involve only two deciles. Of particular interest are the logarithms of ratios of each decile to the median, that is

$$r_t(j) = d_t(j) - d_t(5) = \log\{D_t(j)/D_t(5)\} = \log\{R(j)\}, \quad j = 1, .., 4, 6, .., 9; \quad (5.2)$$

see, for example, Cowell (1995, p29). Of course, the price deflator is now irrelevant. Figure 4 shows the trend extracted from $r_t(1)$ for males using a smooth trend BSM. The antilogarithm has been taken so it is in ratio form, in other words an estimate of $R_t(1)$. It can be seen that the increasing inequality up to 1987 is reversed thereafter, albeit with a slight blip in the early 90s. The fit provided by the smooth trend BSM was relatively better for the decile ratios as compared with the raw deciles. The serial correlation was lower and in many cases the smooth trend constraint, $\sigma_\eta^2 = 0$, could not be rejected with a LR test. The seasonal component usually came out as time-varying but very small with the χ_3^2 test at the end of the sample being statistically insignificant at conventional levels. However when it was dropped the serial correlation increased at lags 2 and 4 and so it seems better to retain it. Table 4 shows the diagnostic and goodness of fit statistics corresponding to those in table 2. Note that the q_ζ ratios are almost identical; the implied weighting pattern essentially corresponds to that in panel (ii) of figure 2.

The decile ratios, or equivalently the deciles themselves, could be combined as follows to give an overall measure of inequality

$$I_t = \sum_{j>5} r_t(j) - \sum_{j<5} r_t(j) = \sum_{j>5} d_t(j) - \sum_{j<5} d_t(j). \quad (5.3)$$

If the distribution is diverging this contrast will be increasing over time. It can be re-arranged in various ways which contrast deciles above and below the median. For example,

$$I_t = \{d_t(9) - d_t(1)\} + \{d_t(8) - d_t(2)\} + \{d_t(7) - d_t(3)\} + \{d_t(6) - d_t(4)\}.$$

For some purposes it may be useful to look at the contrasts above and below the median, that is the two parts of (5.3). It may also be useful to compare the deciles near the median with the highest and lowest deciles, for example by contrasting $r_t(1)$ with $r_t(4)$.

5.1. Stability tests

Suppose we wish to test whether a particular contrast is stable. This then amounts to testing the null hypothesis that, apart from a possibly evolving seasonal component, the contrast series is stationary, with no time trend, against the alternative

that it contains a nonstationary trend component. Thus the relevant test statistic is η, not η_2. With the deciles themselves following non-stationary processes this means that we are testing whether $\boldsymbol{\alpha}$ is a co-integrating vector.

An example of the stability test is provided by $r_t(1)$, which was plotted in figure 4. The test against a random walk component, modified to allow for seasonality[7], is 1.79 which shows a clear rejection as the 5% critical value is 0.46.

Section 6 considers joint tests of several contrasts, including a test of the stability of the whole distribution.

5.2. Divergence and convergence tests

If the contrast is not stable we may wish to put a confidence interval around the underlying change which has taken place. This requires we find the root mean square error (RMSE) of the estimator of $\mu_t - \mu_1$ with particular interest on the case of $t = T$. (The RMSE of an estimator of a stochastic quantity is the same as the variance of the estimation error). We will call combinations involving different time periods 'temporal (level) contrasts'.

A temporal level contrast may be estimated by adding it to the state vector. For a model containing a general stochastic trend of the form (3.3), let $\mu_t^* = \mu_t - \mu_1$ and add the following transition equation

$$\mu_t^* = \mu_{t-1}^* + \beta_{t-1} + \eta_t, \qquad t = 2, .., T. \tag{5.4}$$

The initial value is $\mu_1^* = 0$ with a MSE of zero. The contrast is then tracked by the Kalman filter so its evolution can be studied. The estimate of μ_T^*, together with its RMSE, is obtained when all the observations have been processed. A test, possibly one-sided, may then be carried out on the level contrast statistic,

$$\tilde{\mu}_T^* / RMSE(\tilde{\mu}_T^*), \tag{5.5}$$

based on the standard normal distribution. Of course contrasts can be computed over shorter periods by starting (5.4) in a later time period.

If the series is relatively long, the test statistic (5.5) can be computed from the KFS without adding the extra transition equation. We have

$$\tilde{\mu}_T^* = \tilde{\mu}_T - \tilde{\mu}_{1|T},$$

---

[7]The test was carried out on the seasonally adjusted series obtained by fitting an unrestricted BSM; see Harvey and Streibel (1997) for a justification.

where $\tilde{\mu}_{1|T}$ is the smoothed estimator of μ_1, and, since the correlation between the estimators of μ_1 and μ_T will be negligible,

$$MSE\left(\tilde{\mu}_T^*\right) \simeq MSE\left(\tilde{\mu}_T\right) + MSE\left(\tilde{\mu}_{1|T}\right) = 2MSE\left(\tilde{\mu}_T\right). \qquad (5.6)$$

A nonparametric test, constructed to have power against the alternative of a deterministic but monotonically increasing (decreasing) trend, was suggested by Brillinger (1989). He considers the linear combination $\Sigma w_t y_t$ where

$$w_t = \left\{(t-1)\left(1 - \frac{(t-1)}{T}\right)\right\}^{\frac{1}{2}} - \left\{t\left(1 - \frac{t}{T}\right)\right\}^{\frac{1}{2}}$$

The weighting pattern means that it strongly contrasts the beginning and end levels of the data. The test statistic

$$\tau = \Sigma w_t y_t \Big/ \left\{\hat{\sigma}_L^2 \Sigma w_t^2\right\}^{\frac{1}{2}}, \qquad (5.7)$$

where σ_L^2 is an estimator of the long-run variance as used in the KPSS test, has a standard normal distribution when the level is constant. The test is one-sided on the positive (negative) tail. The test will, of course, have power in any situation where the beginning and end of the series are at a different level.

Continuing the example of $r_t(1)$ at the end of the last sub-section, the STAMP output shows that the level at the end of the series is -0.707 with a RMSE of 0.007. The level at the beginning is -0.699 so $\tilde{\mu}_T^* = 0.008$. This is not significantly different from zero since the RMSE computed from (5.6) is .010 leading to a level contrast statistic, (5.5) of 0.8. On the other hand the level in 87q1 is -0.823 and the difference between this and the level at the end is 0.116 which is clearly significant.

5.3. To log or not to log?

From the point of view of estimating trends, the decision as to whether to take logarithms depends primarily on goodness of fit and diagnostic criteria. If a model is fitted in logarithms, antilogarithms may be taken to convert to levels or ratios, as was done in figure 4. However, taking logarithms does have implications for tests of stability and divergence. If there are upward or downward movements in the deciles, a constant decile ratio does not imply a constant absolute difference. Thus $D_t(j) - D_t(5)$ could be increasing while $D_t(j)/D_t(5)$, and hence $r_t(j)$, is

13

constant. As a result we may wish to test hypotheses regarding the absolute differences as well as, or instead of, ratios.

Logarithms also have implications for forecasting. However, the extrapolation of current trends in the present context is particularly hazardous, and the models were not fitted with this in mind. The aim is simply to extract information concerning past movements and the direction in which various deciles and contrasts appear to be heading in the short term. Indeed long-run predictions would be nonsense since, with different slopes, the deciles would eventually cross. If one really wanted to make coherent long-run predictions it would be necessary to look at variables like the logarithm of the (positive) difference in two deciles, since taking antilogarithms would then give a lower bound of zero.

6. Multivariate models

The local linear trend model can be generalised to the multivariate case straightforwardly simply by writing

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \qquad t = 1, \ldots, T, \qquad (6.1)$$
$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta),$$
$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\zeta), \qquad (6.2)$$

where $\mathbf{y}_t$ is an $N \times 1$ vector and $\boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\zeta$ are $N \times N$ covariance matrices. Seasonals can be added to give a multivariate BSM.

With nine series of deciles there are 45 parameters to estimate in each covariance matrix. The covariance matrices may be restricted in various ways which may lead to statistical gains and/or yield an interpretation. Possibilities include the following.

(i) Correlations decreasing as deciles are further apart. For example, one would expect the first and second deciles to move more closely together than the first and ninth.

(ii) Variances (diagonals) the same.

(iii) Matrices proportional, that is $\boldsymbol{\Sigma}_\eta = q\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_\zeta = q_\zeta \boldsymbol{\Sigma}_\varepsilon$. This is known as homogeneity. Such models are very easy to estimate; see Harvey (1989, ch 8).

(iv) The matrices $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\zeta$ being of reduced rank. In particular if the wage distribution were stable over time, there would be a single common trend and the

14

model could be written:

$$\mathbf{y}_t = \mathbf{i}\mu_t + \overline{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_t, \tag{6.3}$$

where μ_t is a univariate stochastic trend and $\overline{\boldsymbol{\mu}}$ is a vector of constants, one of which is set to zero for reasons of identifiability (unless μ_0 is set to zero). It is possible that part of the distribution is stable over time so there are several common trends associated with groups of deciles. Estimating the full model and conducting a principal components analysis on the estimate of $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\zeta$ may give an indication of possible structures.

A multivariate model may enable trends to be estimated more efficiently. The gain depends on the relationship between the covariance matrices $\boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\zeta$. For example if they are proportional (homogeneity) there is no reduction in the MSE of the estimated trend, though the precision with which the common signal-noise ratios may be estimated increases.

If the observations are modelled by a multivariate BSM, then any set of linear combinations will satisfy a model of the same form. If $\mathbf{A}$ is an $R \times N$ matrix, the covariance matrices in the multivariate BSM for $\mathbf{A}\mathbf{y}_t$ will be of the form $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$. Note that a homogeneous model will remain homogeneous.

6.1. Multivariate stationarity tests

In the multivariate local level model, that is (6.1) without the slope, a test of the null hypothesis that $\boldsymbol{\Sigma}_\eta = \mathbf{0}$ is a test against the alternative that there is nonstationarity in the system. Nyblom and Harvey (2000b) show that an LBI test can be developed against the homogeneous alternative $\boldsymbol{\Sigma}_\eta = q\boldsymbol{\Sigma}_\varepsilon$. The test has the rejection region

$$\eta(N) = \text{tr}\left[\mathbf{S}^{-1}\mathbf{C}\right] > c, \tag{6.4}$$

where

$$\mathbf{C} = T^{-2}\sum_{i=1}^{T}\left[\sum_{t=1}^{i}\mathbf{e}_t\right]\left[\sum_{t=1}^{i}\mathbf{e}_t\right]' \quad \text{and} \quad \mathbf{S} = T^{-1}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t'. \tag{6.5}$$

where $\mathbf{e}_t = \mathbf{y}_t - \overline{\mathbf{y}}$. Under the null hypothesis, the limiting distribution of (6.4) is the Cramér-von Mises distribution with N degrees of freedom, denoted $CvM(N)$.

Although the test maximizes the power against homogeneous alternatives, it is consistent against all nonnull $\boldsymbol{\Sigma}_\eta's$. This follows from the result that $T^{-1}\eta(N)$ has a nondegenerate limiting distribution. In fact the limiting distribution depends only on the rank of $\boldsymbol{\Sigma}_\eta$. This result suggests that an indication of the number of common trends can be obtained, without estimating a model, simply by looking

at the eigenvalues of $\mathbf{S}^{-1}\mathbf{C}$. A formal test of the null hypothesis that there is a specific number of common trends, that is

$$H_0 : rank(\mathbf{\Sigma}_\eta) = K \qquad \text{against} \quad H_1 : rank(\mathbf{\Sigma}_\eta) > K, \quad K < N,$$

can be based on the sum of the $N - K$ smallest eigenvalues of $\mathbf{S}^{-1}\mathbf{C}$; see Nyblom and Harvey (2000b).

6.2. A test of stability

If the distribution is stable over time, as in (6.3), there are eight co-integrating vectors. Thus we need to choose eight contrasts with the property that $\mathbf{Ai} = \mathbf{0}$ and test whether they are jointly stationary. As noted earlier a standard way of setting up contrasts is with respect to the median as in (5.2). In this case $\mathbf{A}$ has $R = (N - 1)$ rows each with minus one in the fifth position and one in position j for $j \neq 5$.

The test of stability is carried out by applying the test of (6.4) to $\mathbf{y}_t = \mathbf{Ad}_t$. As observed in sub-section 5.1, time trends would not normally be present under the null and this is also apparent from (6.3). The test statistic

$$\eta(N - 1; \mathbf{A}) = \mathrm{tr}\left[(\mathbf{ASA}')^{-1}\mathbf{ACA}'\right] \qquad (6.6)$$

is therefore formed from deviations from the means. Its limiting distribution under the null hypothesis is $CvM(N - 1)$.

Other choices of $\mathbf{A}$ are possible since if the original $\mathbf{A}$ is pre-multiplied by a non-singular $R \times R$ matrix the test statistic is unchanged. Note that in general the $R \times N$ matrix $\mathbf{A}$ must be of rank R.

Several contrasts could be tested separately. For example we may wish to test if there is partial stability among the decile ratios, in other words a common trend for some of them.

Note that a homogeneous model cannot have common stochastic trends, so it cannot have stable contrasts.

6.3. Results

Attempting to fit all nine series was not satisfactory and in fact a higher likelihood was achieved by imposing the homogeneity constraint. For a better comparison we fitted the odd deciles only. The program converged and the model gave a satisfactory fit. Figure 5 plots the trends indexed to zero in 1979q1 so they are

all on the same scale as the first decile. The homogeneous model just failed to be rejected on a likelihood ratio test at the 5% level of significance. The test statistic was 22.84, while the 5% critical value for a χ^2 with $N(N+1)/2 - 1 = 14$ degrees of freedom is 23.69. The summary statistics for the five deciles are similar to those given for the univariate models in table 2. The (common) estimate of q_ζ was 0.25. The five estimates for the irregular variance, σ_ε^2, are (when multiplied by 10^5) 6.4, 6.1, 6.3, 5.4, 7.7. One would probably be reluctant to impose a constraint that they are all the same.

Table 4 shows the correlations implied by the homogeneous covariance matrices. The correlation tends to be higher for deciles close together.

The four decile ratios, $r_t(j), j = 1, 3, 7, 9$ were estimated successfully by an unrestricted smooth trend BSM. The antilogs of the trends are shown in figure 6. The trend in panel (i) is the same as in figure 4 and it can be seen that the stochastic movements in the other trends are of a similar order of magnitude - it will be remembered that the hypothesis of a constant level for $r_t(1)$ was overwhelmingly rejected. It can be seen that before 1985 inequality was increasing all round in that the lower decile ratios, $r_t(1)$ and $r_t(3)$ were moving downwards while the upper ones were moving upwards. In the late 1980s all the ratios were moving up. This pattern continues again after the early 90s. Thus the lower deciles are getting closer to the median while the higher ones are moving further away. We discuss some analytic implications of this shift in the pattern of wage inequality in the conclusion.

In view of the results in table 4, it is not surprising that the homogeneous model cannot be rejected at the 5% level of significance though it can be at the 10%; the LR statistic is 16.33.

7. Male and female wages

Figure 7, top panel, shows the ratio of female to male median wage obtained by fitting a smooth trend model to the logarithm of the ratio and taking antilogarithms. A cyclical component was included, rather than just working with the BSM. This gave a smoother trend, the salient feature of which is the steady relative increase in female wages up to 1993, after which it levels off. Examining the trends of the two variables comprising the contrast, we find that the 13.7% decline in the trend of the real male median wage between 1979q1 and 1993q1 underlies the relative gains of females; the trend in their real wage grew by 5.9% over this period. Since 1993q1, the trend in the gender wage gap has been quite flat, with

the trend growth of both male and female real median wages being about 4%. These findings are confirmed formally by the temporal level contrast statistics as applied to the logarithm of the trend, which at the of the period was -0.250 with a RMSE of 0.0088. Formula (5.6) gives RMSE($\widetilde{\mu}_T^*$) as 0.012 and over the full period (5.5) is 17.9 which is highly significant. The flatness of the gender gap towards the end is confirmed by noting that in mid-1993 the trend was -0.257. Even though (5.6) may be a less satisfactory approximation in this case, it is clear that there is no significant change.

The short-term movements in the cycle, shown in the lower panel of figure 7, are somewhat irregular in period and amplitude, although there is some evidence of spikes at five year intervals. These intervals do not, however, correspond to any of the cyclical indicators, such as gender-specific unemployment rates, which might be expected to play a role in the gender gap. There are many other non-cyclical factors that might be called upon to explain the compressing of the gender gap throughout the 1980s. As documented by Blau (1998), women made consistent progress over this period on some key wage determinants, primarily labor market experience and occupational upgrading. Variables like the average experience level of the female workforce or the share of female white-collar workers could be brought into the BSM. This could help shed light on the role of these factors in both the closing of the gender gap in the 1980s and its subsequent (and as yet, unexplained) deceleration in the 1990s.

8. Minimum wage

The US Federal minimum wage is a binding wage floor on the wage distribution, covering about 90% of the wage and salary workforce.[8] It is typically argued that the minimum wage plays an important wage setting role for low-wage workers in the US; see, for example, Spriggs and Klien (1994). Others, such as DiNardo *et al.* (1996), have emphasized the role of the long-term decline in the real minimum wage in explaining the increase in wage inequality over the 1980s, when the minimum fell 30% in real terms.

Figure 8 shows the lowest three female real wage deciles, together with the real minimum wage. (We prefer to work here in actual real values rather than

---

[8]This share refers to those workers subject to the Fair Labor Standards Act Minimum, Wage Provisions (U.S. DOL, 1998). States can set their minimums above or below the federal level, but they cannot pay covered workers less than the federal minimum. In 1999, nine states had minimum wages above the federal level.

logarithms for the purposes of interpretation). At the beginning of the series the first decile, $F_t(1)$, is below the minimum wage and the increase in the minimum wage in 81q1 leads to a corresponding large and instantaneous increase in the series. (The rise in 80q1 has very little impact).[9] The minimum wage and $F_t(1)$ stay close for a few years and then in the mid-eighties, the minimum wage starts to fall way below $F_t(1)$ until at the beginning of the nineties the gap is about 85 cents. The increase in the minimum wage in 90q2 has minimal impact, serving only to bring it back in line with $F_t(1)$. Similarly the minimum wage rises of 1996q4 and 1997q4 have little effect on $F_t(1)$, since the increase in $F_t(1)$ in the second half of the nineties is also to be found in $F_t(3)$ which, as we establish shortly, is unaffected by minimum wage increases. These findings are consistent with those of Card (1992), who shows that the impact of a minimum wage increase is greater when it has more "bite," in other words it affects a larger share of the low-wage workforce.

Fitting the smooth trend BSM to $F_t(1)$ gives a Bowman-Shenton normality test statistic equal to 556.1; the 5% critical value is a mere 5.99. This is a reflection of the enormous jump in response to the minimum wage increase of 81q1. Putting in a level dummy variable[10] at this point reduces the normality statistic to 6.61, which, although significant at the 5% level, is quite tolerable; see table 6. The dummy is highly significant with a 't-statistic' of 13.14 and the model as a whole fits well with diagnostics, such as Box-Ljung, being perfectly acceptable and the trend, shown in figure 9, providing a good indication of the underlying movements in the series. The dummy is incorporated into the trend, so the level increases in 81q1 by the dummy's coefficient of 0.52. Table 6 also shows results of fitting the same model to other deciles. As can be seen, there is a minimum wage spillover effect onto $F_t(2)$, where the increase in 81q1 is 0.23. However, the third decile is apparently unaffected, as is the first male decile, $M_t(1)$.

9. Conclusion

This article has shown how unobserved components time series models may be used to exhibit the stylized facts of time series of wage percentiles. In particular the trends in various measures of inequality based on combinations of percentiles

---

[9]This may be due to the fact that inflation was quite a bit higher than it was a year later; thus, the 80q1 real increase was 3.6% while the 81q1 real increase was 5.3%.

[10]This is done by adding λw_t to the first equation in (3.3), with $w_t = 1$ in 81q1 and zero thereafter. The level shift is then incorporated into the trend.

may be tracked over time. Tests of stability and divergence (or convergence) are derived using recently established time series theory on co-integration and common trends. Multivariate models can be fitted to groups of percentiles and the correlations of the various components across groups studied.

The methods were applied to a newly constructed data set of quarterly wage deciles derived from the outgoing rotation groups of the Current Population Survey. The main substantive conclusions are that the character of hourly wage inequality shifted in the mid-1980s, from one of generalized divergence throughout the wage scale, to one where the top deciles pulled away from those in the middle and bottom of the wage scale (this shift occurred toward the end of the 1980s for female workers). Thus, $r_t(9)$–the logarithm of the 9th decile minus the that of the median–grew consistently throughout the full period for both sexes while that of $r_t(1)$ flattened and even increased (implying less inequality) in the 1990s.

These facts need to be addressed by those who propose explanations for the increase in US wage inequality. For example, one of the more popular explanations argues that skill-biased technological change (SBTC), brought about by the increased use of computers in the workplace, is a primary cause of growing wage differentials. Yet, the trends shown above raise the question as to what type of technological process would generate this observed pattern, differentiating the deciles throughout the wage scale in one period, and only those at the top of the wage scale in another period. Is it plausible that SBTC has persistently lifted the relative wages of high wage workers, but only did so for middle-wage workers through the mid-1980s? Further research which accounts for these nuances in the pattern of wage inequality is necessary.

Similarly, the trend in the so-called gender gap, which we measure as the difference between female and male median wages, levelled off in the 1990s, after steadily closing throughout the 1980s. Here too more research is needed to identify the factors that explain the trend's flattening in the 1990s. One approach would be to add explanatory variables, such as the average level of female labor market experience or occupational upgrading (e.g., the share of female white collar workers), to the BSM of the gender wage gap and examine whether either these variables themselves or their relative returns diminished in the 1990s.

As regards the minimum wage, the only increase which had a measurable impact on the lowest female decile was the one in 81q1. This happened because the minimum wage was actually higher than the lowest female decile at the time. Its effect spread over into the second decile but not the third. The male deciles

20

were apparently unaffected. Later increases in the minimum wage have little impact on the first female decile as they are some way below its level.

ACKNOWLEDGMENTS

Table 1 Estimated models for the logarithm of median male wages

	Trend Specification			
	Unrestricted	RWD	RW	IRW
s.e.	.0109	.0106	.0107	.0111
Q(6)	5.60	7.27	7.69	7.46
r(1)	-.00	-.03	-.03	.14
log L	343.66	343.90	349.54*	341.51

Table 2. Univariate smooth trend models for logs of male deciles

	Deciles				
Statistic	1	3	5	7	9
q_ζ	.22	.29	.28	.13	.26
SE	.0111	.0107	.0111	.0099	.0121
$Q(8,6)$	6.02	9.83	7.46	25.46	9.69
$r(1)$	.06	.15	.14	.22	.12

Table 3. Seasonal patterns for male deciles

	Deciles				
Quarter	1	3	5	7	9
1	.23	.64	.71	.55	.86
2	-.15	-.11	-.11	-.13	-.21
3	-.49	-.83	-.86	.69	-.89
4	.42	.31	.27	-.27	.23

Table 4. Univariate smooth trend models for decile ratios

	Contrasts			
	1	3	7	9
q_ζ	.26	.25	.26	.26
SE	.0140	.0098	.0094	.0142
$Q(8,6)$	7.73	8.33	5.24	12.53
$r(1)$	.05	.08	.14	.09

Table 5. Correlations between deciles in homogeneous model for male log deciles

	Decile				
Decile	1	3	5	7	9
1	1	.49	.22	.12	.16
3		1	.62	.38	.23
5			1	.61	.21
7				1	.41
9					1

.

Table 6. Models for lower male and female deciles with level dummy at 1981q1.

Decile	Dummy	$t-ratio$	N	$Q(8,6)$	q_ζ
F(1)	.52	13.14	6.62	7.29	.70
F(2)	.23	4.82	9.21	3.65	.29
F(3)	-.04	-.70	4.36	9.21	.24
M(1)	.05	.90	.61	5.13	.22

10. References

Bernstein, Jared and Lawrence Mishel, "Has Wage Inequality Stopped Growing?" *Monthly Labor Review* 120:12 (1997), 3-16.

Blau, Francine, "Trends in the Well-Being of American Women, 1970-95," *Journal of Economic Literature*, 36:1 (1998), 112-165.

Brillinger, D.R. "Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series," *Biometrika,* 76 (1989)**,** 23-30.

Canova, F. and B.E. Hansen (1995), "Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability," *Journal of Business and Economic Statistics,* 13**,** 237-52.

Card, David, "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage." *Industrial and Labor Relations Review*, 46 (1992), 22-37.

Committee on Ways and Means, *The Green Book: Overview of Entitlement Programs.* U.S. Government Printing Office, Washington, DC (1992), 1358.

Cowell, F A, *Measuring Inequality.* Harvester Wheatsheaf: Hemel Hempstead (1995).

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 65 (1996), 1001-44.

Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press (1989).

Harvey, A.C. and Jaeger, A., "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, 8 (1993), 231-47.

Harvey, A. C. and Koopman, S.J., "Signal Extraction and the Formulation of Unobserved Components Models," CentER discussion paper 9944 (1999), Tilburg University.

Harvey, A.C. and M. Streibel, "Testing for nonstationary unobserved components," mimeo (1997)

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N., *STAMP 5.0 Structural Time Series Analysis Modeller and Predictor*, London: Chapman and Hall (1995).

Katz, Lawrence and David Autor, "Changes in the Wage Structure and Earnings Inequality," in Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics,* Amsterdam, North Holland (1999).

Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Y.Shin , "Testing the Null

Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?" *Journal of Econometrics* 44, 159-78 (1992).

Mishel, Lawrence, Jared Bernstein, and John Schmitt. *The State of Working America*, 1998-99. Ithaca, NY: Cornell University Press (1999).

Nyblom, J. and T. Mäkeläinen, "Comparison of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association, 78, 856-64 (1983).*

Nyblom, J., and A.C.Harvey, "Testing against smooth stochastic trends," mimeo (2000a).

Nyblom, J., and A.C.Harvey, "Tests of Common Stochastic Trends, *Econometric Theory,* 16 (to appear, 2000b)

Pfeffermann, D., "Estimation and Seasonal Adjustment of Population Means Using Data from Repeated Surveys," *Journal of Business and Economic Statistics*, 9, 163-75 (1991).

Polivka, Anne E., "Using Earnings Data for the Current Population Survey After the Redesign," Working Paper #306, U.S. Bureau of Labor Statistics, Washington, DC (1998).

Spriggs, William and Bruce W. Klein, "*Raising the Floor: The Effects of the Minimum Wage on Low-wage Workers.*" Washington, D.C.: The Economic Policy Institute (1994).

U.S. Department of Labor, "Minimum Wage and Overtime Hours Under the Fair Labor Standards Act," Employment Standards Administration, Wage and Hour Division (1998).

West, Sandra A., "Measures of Central Tendency for Censored Earnings Data from the Current Population Survey," unpublished Bureau of Labor Statistics report (undated).