PRB 7|24|02
MOELLER
2:30 PM

Accounting for Actual Work Experience:

Prototype Hedonic Labor Quality Indexes

TO BE PRESENTED AT THE NBER SUMMER INSTITUTE 2002

PRODUCTIVITY POTPOURRI WORKSHOP

PRELIMINARY DRAFT

NOT FOR CITATION OR QUOTATION

Linda Moeller

Office of Productivity and Technology

Bureau of Labor Statistics

Moeller_L@bls.gov

July 17, 2002

This paper reports on research undertaken by Linda Moeller under an inter-agency agreement between the Bureau of Labor Statistics and the Bureau of the Census. The results and conclusions discussed here are those of the author and do not represent the views of the Office of Productivity of the Bureau of Labor Statistics (BLS), Department of Labor, or the Bureau of the Census (BOC), Department of Commerce. The results are preliminary and should not be quoted or cited. This draft is being released to inform interested parties of this research and to encourage discussion.

Accounting for Actual Work Experience:

Prototype of Labor Quality Hedonic Indexes

Table of Contents

Introduction

The BLS multifactor productivity series decompose labor productivity growth into components associated with increased capital intensity of production and changes in the skill-composition of the work force. The labor composition index, which serves to gauge the latter effect, is constructed with hedonic techniques motivated by classical human capital theory. More specifically, the labor composition index is constructed by aggregating within-cell growth in hours worked by persons with varying levels of schooling and experience, using weights constructed with predicted values from a human capital wage equation.

The use of administrative record data on actual accumulated work experience as an indicator of workers' current productivity is a distinguishing features of this labor composition series. The current procedure relies on an experience proxy that is based on a one-time match of Social Security Administration (SSA) data to records from the 1973 March CPS Income Supplement. Since the parametric relationship between accumulated work experience and the demographic characteristics of the work force is unlikely to have remained stable over time, the BLS has undertaken a long-run research project to update the work experience data at regular intervals. This paper reports on research into the construction of a prototype index constructed with longitudinal microdata, including retrospective information on accumulated total work experience, from the Survey of Income and Program Participation (SIPP).

Prior research with SSA administrative record data linked to the 1984 panel of the SIPP revealed that SSA-based estimates were biased downward in the case of older workers, due to incomplete Social Security coverage during the initial years of that program.[1] Social Security coverage rates have increased over time, and SSA records now include most of the work history of most of the current workforce. However confidentiality concerns may preclude the routine linking of household survey data and administrative records on an annual basis. Therefore subsequent research has focused on overlapping panels of the SIPP that span the period 1984-1993.[2] In this research, modification of the current wage equation estimation procedure to incorporate an inverse Mills ratio selection bias correction factor proved to be straightforward.

It is a strong suit of this index that it is based on large, well-designed nationally representative surveys that collect a large number of variables of interest to labor economists.[3] The specification and estimation of the wage equation follows a conventional structural approach. A structural approach is attractive in that it supports the quantitative analysis of the underlying economic forces that cause changes in the index over time, and an attempt is made to achieve greater consistency with the other components of the official BLS productivity series by allowing for industry wage

---

[1] Linda L. Moeller (2002), "On the Estimation of Classical Human Capital Wage Equations with Two Independent Sources of Data on Actual Work Experience," BLS Working Paper 362, available at http://stats.bls.gov/ore/pdf/ec020110.pdf

[2] Moeller (1999), "A Second Decade of Slower U.S. Productivity Growth: Prototype Labor Composition Indexes Based on the SIPP," presented at the 1999 NBER Summer Institute Workshop on Price, Quantity and Quality Measurement and available at http://www.nber.org/~confer/99/prbgsi99/moeller.pdf.

[3] Thus this research project is in the spirit of earlier work by Newey, Powell and Walker, who found the specification of the estimating equations to be of greater importance than semiparametric estimation techniques in applied work. See Whitney K. Newey, James L. Powell and James R. Walker (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results," American Economic Review, Papers and Proceedings, May, pp. 324-328.

differentials.[4] This research project may be contrasted with current research by Abowd, Haltiwanger and Lane et. al. that is based on very large sets of administrative records but a small number of variables.[5] It may also be contrasted with the research of Jorgenson, Gollop and Fraumeni in which labor composition indexes are constructed with panels of time series aggregates.[6]

The estimated coefficients of three-equation systems of equations are intermediate products in the construction of the labor composition index. The endogenous variables of these equations are actual accumulated work experience, the probability of employment during the current time period, and observed wage rates. Wage rates are estimated as functions of workers' educational achievement and accumulated work experience, and broadly-defined categorical variables for primary industry of employment that capture long-run differences in the capital intensity of production as well as cyclical variation in the demand for labor services. The wage equation incorporates a Heckman-type selection bias correction factor based on the estimated coefficients of a probit equation in which household characteristics and regional categorical variables determine the difference between offered and reservation wage rates, and thus the probability of employment. Since the employment rate is the dependent variable of the probit, the

---

[4] For a discussion of the identification problem as it arises in the estimation of hedonic equations, see Shulamit Kahn and Kevin Lang (1988), "Efficient Estimation of Structural Hedonic Systems," International Economic Review, 29, pp. 157-166, Kenneth G. Stewart and J.C.H. Jones (1998), "Hedonics and Demand Analysis: The Implicit Demand for Player Attributes," Economic Inquiry, pp. 192-202, and the references cited in those articles. In the current application, the use of categorical variables for industry and occupation to control for industry-specific differences in human and physical capital employed in production may correspond roughly to Kahn and Lang's discussion of variables that vary with the "matching process" in the context of multiple markets.

[5] See, for example, John M. Abowd, John Haltiwanger, Ron Jarmin, Julia I. Lane, Paul Lengermann, Kristin McCue, Kevin McKinney and Kristin Sandusky (2002), "The Relationship Between Human Capital, Productivity and Market Value: Building Up From Micro Evidence," photocopy.

regional categorical variables should capture regional differences in labor market slackness.

The focus of the current paper is on the specification and estimation of the work experience equation, and the incorporation of a predicted work experience variable into selection-bias corrected wage equations. This work will be used, in part, to evaluate the advisability of adopting two-sample instrumental variable (TSIV) procedures to construct the official BLS labor composition index.[7] Specifically, in future work the coefficients of a work experience equation estimated with microdata from the SIPP may be applied to annual data from the March Current Population Survey (CPS), to construct a proxy for actual work experience.[8] This approach is feasible because the CPS and the SIPP share a common sampling frame, and have many key variables in common.

Data Description

The current BLS labor composition index is constructed primarily with microdata from the annual March supplement to the CPS. In general, the CPS is designed to calculate employment and unemployment rates on a monthly basis. The SIPP is focused on

[6] Dale W. Jorgenson, Frank M. Gollop and Barbara M. Fraumeni (1987), Productivity and U.S. Economic Growth, Cambridge, Mass.: Harvard University Press.

[7] Recent work in this area includes Joshua D. Angrist and Alan B. Krueger (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," Journal of the American Statistical Association, 87 (418), pp. 328-336; Angrist and Krueger (1995), "Split-Sample Instrumental Variables Estimates of the Returns to Schooling," Journal of Business and Economic Statistics, 13 (2), pp. 225-235; and Phoebus J. Dhrymes and Adriana Lleras Muney (2001), "Estimation of Models with Grouped and Ungrouped Data by Means of '2SLS,' available at http://www.princeton.edu/~alleras.

[8] The SIPP has been conducted since 1984. For a good introductory overview see Thomas B. Jabine, with Karen E. King and Rita J. Petroni (1990), SIPP Quality Profile, Washington: Bureau of the Census, Department of Commerce

income and program participation. The two surveys are designed to function as interlocking components within a system of household surveys that also includes the Health and Retirement Survey and the American Community Survey. The SIPP and the CPS are drawn from the same sampling frames, SIPP sampling weights are benchmarked to CPS population estimates, and both surveys collect information on the employment and earnings of households. The two surveys exhibit complementary strengths. The CPS sample size is larger and CPS data are available on a more timely basis. The SIPP is a longitudinal survey that collects more detailed retrospective information, including information about respondents' past work histories.

These complementarities between the CPS and the SIPP suggest that proxies constructed through TSIV estimation procedures should be strong, rather than weak. Because the number of variables collected is quite large, these surveys support a more structural approach to model specification, and thus more efficient parameter estimates, than is generally possible in applications that involve much larger samples of administrative records. Consequently, as noted above, an effort has been made to pursue the analytical insights available through the specification and estimation of identified linear systems of wage and experience equations for estimation with two-stage least squares (TSLS) procedures, and to adhere fairly closely to wage equation specifications that are generally accepted within the literature.

Description of Labor Composition Index

The labor composition index, denoted $\frac{\dot{C}}{C}$ in the discussion that follows, is a component of the BLS multifactor productivity series that is intended to reflect changes in the skill composition, or the human capital intensity, of the work force. The multifactor productivity series measures the contribution of increased capital intensity to productivity growth, and the labor composition index is a parallel estimate of the contribution of increases in the human capital of the work force to multifactor productivity growth.

More formally, multifactor productivity growth (MFP) is defined as the difference between the growth of output and the growth of a weighted sum of inputs, assuming that the inputs are paid the value of their marginal products on average. Imposing separability assumptions under which an aggregate production of the form $Q_t = A_t F(K_t, L_t)$ may be said to "exist," logarithmic derivatives of the production function are rearranged to obtain an expression for MFP, or $\frac{\dot{A}_t}{A_t}$ in the expression below:

$$MFP \equiv \frac{\dot{A}_t}{A_t} = \frac{\dot{Q}_t}{Q_t} - s_K \frac{\dot{K}_t}{K_t} - s_L \frac{\dot{L}_t}{L_t}.$$

The shares of capital and labor respectively in national income are denoted s_K, s_L. K_t and L_t are value-weighted stocks of capital and labor employed at time t.[9]

[9] This exposition is generally well known. See, for example, Charles R. Hulten (2001), "Total Factor Productivity: A Short Biography," in New Developments in Productivity Analysis, edited by Charles R.

Multifactor productivity growth rates are an estimate of the degree to which labor productivity growth is attributable to shifts in the composition of growth in physical assets employed in production, and the labor composition index is a parallel estimate of the degree to which labor productivity and multifactor productivity growth rates are attributable to shifts in the skill-composition of growth in hours worked. More specifically the labor composition index serves to decompose growth in labor services, or $\frac{\dot{L}_t}{L_t}$, into growth in hours worked, or $\frac{\dot{H}_t}{H_t}$, and changes in the skill-composition of the work force $\frac{\dot{C}_t}{C_t}$.

Labor composition index weights ω_l, used to aggregate hours worked by different categories of workers, are two-year averages of the labor cost shares of workers with characteristics l. Letting $H_{l,t}$ denote hours worked by workers with characteristics l at time t, omitting subscripts from aggregated values, and omitting time subcripts for simplicity:

$$\frac{\dot{L}}{L} = \omega_1 \frac{\dot{H}_1}{H_1} + \cdots + \omega_n \frac{\dot{H}_n}{H_n} = \omega_1\left(\frac{\dot{H}_1}{H_1} - \frac{\dot{H}}{H}\right) + \cdots + \omega_1\left(\frac{\dot{H}_n}{H_n} - \frac{\dot{H}}{H}\right) + \frac{\dot{H}}{H} \equiv \frac{\dot{C}}{C} + \frac{\dot{H}}{H}.$$

Thus the labor composition effect is $\frac{\dot{C}}{C} \equiv \frac{\dot{L}}{L} - \frac{\dot{H}}{H}$.[10]

Hulten, Edwin R. Dean and Michael J. Harper, NBER and CRIW Studies in Income and Wealth, Vol. 63, Chicago: University of Chicago Press, pp. 1-53.

[10] The exact calculations are described more fully in Moeller (2002 and 1999).

Following a standard methodology for the calculation of a hedonic quality index, BLS uses a standard human capital wage equation to estimate conditional mean wage rates, $\hat{w}_{l,t}$, which are used in turn to calculate labor cost share weights $\omega_{l,t}$ for persons with characteristics l at time t:[11]

$$\omega_{l,t} = \frac{1}{2}\left[\left(\hat{w}_{l,t}H_{l,t} \Big/ \sum_{\forall j,t} \hat{w}_{j,t}H_{j,t}\right) + \left(\hat{w}_{l,t-1}H_{l,t-1} \Big/ \sum_{\forall j,t-1} \hat{w}_{j,t-1}H_{j,t-1}\right)\right].$$

The conditional mean wage rates $\hat{w}_{l,t}$ in the expression above are calculated as predicted values from an OLS regression of the following form.

$$\hat{w}_{l,t} = \hat{\beta}_0 + \hat{\beta}_{s,t}\bar{s}_{l,t} + \hat{\beta}_{e,t}\bar{\tilde{e}}_{l,t} + \hat{\beta}_{d,t}\bar{d}_{l,t} + \hat{\beta}_{r,t}\bar{r}_{l,t}\,.$$

In this last equation overbars denote within-cell weighted sample means. Categorical variables (dummy variables) for the number of years of school completed are represented as $s_{l,t}$, and $\tilde{e}_{l,t}$ is a proxy for years of actual work experience for persons with characteristics l. The demographic variables in the vector $d_{l,t}$ include ever-married

[11] This general approach is discussed extensively in Zvi Griliches (1970), "Notes on the Role of Education in Production Functions and Growth Accounting," in Education, Income and Capital, W. Lee Hanson, Ed., New York: NBER and Columbia University Press, pp. 71-115; and in Griliches (1971), "Introduction: Hedonic Price Indexes Revisited," in Price Indexes and Quality Change: Studies in New Methods of Measurement, Griliches, Ed., Cambridge: Harvard University Press, pp. 3-15. For a contemporary critical discussion of this approach see Ariel Pakes (2002), "A Reconsideration of Hedonic Price Indices with an Application to PC's," NBER Working Paper No. 8715.

status, Black or ethnicity, and full-time/part-time status. Categorical variables in the vector $r_{l,t}$ include region and a city size variable.[12]

Data on actual accumulated work experience are not available in the CPS. It is noteworthy that the current BLS procedure is to use the coefficients of a demographically-driven experience equation, estimated with data from the 1973 CPS that have been linked to SSA employment records, to generate proxies for actual work experience that are entered into wage equations estimated with successive annual cross sections from the March CPS annual income supplement. Regularly-available survey data on accumulated actual work experience may be preferred to linked administrative record data for these purposes.[13]

The Significance of Accumulated Work Experience in Human Capital Theory

The existence of an extensive and longstanding literature on the specification and estimation of human capital wage equations makes the use of hedonic techniques to evaluate changes in the skill composition (or "quality") of labor services particularly attractive. Within this analytical framework the determinants of accumulated

---

[12] The procedures currently used by BLS are described more completely in BLS Bulletin 2426, Labor Composition and Productivity Growth, 1948-90, U. S. Department of Labor, Bureau of Labor Statistics, December 1993.

[13] Econometric procedures that might be implemented with the SSA data, to compensate for selection bias due to relatively low Social Security coverage rates at the inception of the program and time-varying ceilings on recorded earnings, are discussed in Marjorie Honig and Giora Hanoch (1985), " 'True' Age Profiles of Earnings: Adjusting for Censoring and for Period and Cohort Effects," Review of Economics and Statistics, pp. 383-394. However the simplifying assumptions that would be required to adjust for both limitations simultaneously are non-trivial. Also see Claudia Goldin (1989), "Life-Cycle Labor Force Participation of Married Women: Historical Evidence and Implications," Journal of Labor Economics, 7(1), pp. 20-47, especially Figure 1.

employment experience, or $\tilde{e}$ in the wage equation above, include the relative marginal productivities of household members in market and non-market production, and household members' current and expected future income and time constraints over the course of the life cycle.[14]

For example, in the simplified static case of a two-person household consisting of a husband and wife, optimal household utility U_t is achieved through the joint consumption of "commodities" $Z_{i,t}$, i.e., by maximizing the following utility function.

$$U_t = U(Z_{1,t}, \cdots Z_{m,t}),$$

where i identifies a commodity consumed during period t. Purchased goods and services are combined with household members' time to produce these commodities:

$$Z_{i,t} = Z_{i,t}(T_{h,i,t}, T_{w,i,t}, X_{h,i,t} X_{w,i,t}).$$

Abstracting from non-wage income, the household budget constraint is specified as a function of the market wage rates of each household member, their time constraints, row vectors of market goods and services $X_{s,i,t} = (X_{s,i,t,1} \cdots, X_{s,i,t,n})$ consumed by each spouse in the production of commodity i, an associated column vector of market prices $p_{i,t}$, and

[14] The analysis in the next few paragraphs are based on Jacob Mincer's labor economics seminar at Columbia, during the 1984-85 academic year. They appear to follow an "Addendum" in the 1983 reprint of Becker's Human Capital, and to be a slightly more detailed extension of the analysis in Gronau's then-forthcoming chapter in the Handbook of Labor Economics. See Gary S. Becker (1983), Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition, Chicago and NBER: Midway Reprint, pp. 56-71; and Ruben Gronau (1986), "Home Production – A Survey," Handbook of Labor Economics, vol. 1, Ashenfelter and Layard, Eds., Amsterdam: North-Holland, pp. 273-304. The seminal work in this area includes Mincer (1963), "Market Prices, Opportunity Costs, and Income Effects," in Measurement in Economics: Studies in Mathematical Economics in Memory of Yehuda Grunfeld, Christ et. al. Eds, Stanford: Stanford University Pres; Gary S. Becker (1965), "A Theory of the Allocation of Time," Economics Journal, 75, pp. 493-517; and Kelvin Lancaster (1966), "A New Approach to Consumer Theory," Journal of Political Economy, 74, pp. 132-157.

therefore by the "full prices" of commodities, $\pi_{i,t}$. The household's "full budget constraint" takes the following form:

$$w_{h,t}\overline{T}_{h,t} + w_{w,t}\overline{T}_{w,t} = w_{h,t}\sum_i T_{h,i,t} + w_{w,t}\sum_i T_{w,i,t} + \sum_i p_{i,t}\left(X_{h,i,t} + X_{w,i,t}\right),$$

where $\overline{T}_{s,t}$ is the maximum amount of time available to household member s for total market and non-market activity, $w_{h,t}$ is that person's market wage rate, and the full price of commodity i is $\pi_{i,t} = w_{h,t}T_{h,i,t} + w_{w,t}T_{w,i,t} + p_{h,t}\left(X_{h,i,t} + X_{w,i,t}\right)$. Optimization with respect to $Z_{i,t}$ shows the ratios of the marginal utilities of these commodities to be equal to their full prices. Optimization with respect to the inputs $T_{h,i,t}, T_{w,i,t}$ and $X_{s,i,t}$ results in first-order conditions that can be expressed in terms of the marginal productivity of each input in the production of each commodity, and its marginal utility.[15]

Within this context it is straightforward to argue that childrearing is a labor-intensive household production activity, and that it is optimal for one member of a two-adult household to specialize in non-market production even if both adults had acquired identical levels of market-oriented human capital when living independently, in view of the fixed costs associated with childrearing. From this perspective the labor force attachment of women is intermittent because women are more likely than men to specialize in non-market childrearing activities during some portion of their working lives. Thus the presence of young children is expected to decrease the likelihood of

[15] As noted by Pollack and Wachter, in practice the effects of technological change often cannot be identified separately from changes in taste because data on time use are available, and because household production often appears to be joint production, or "multi-tasking," in practice. Identifying technological change in household production separately from changes in tastes with respect to household production is not a primary concern of the current project, however. See Robert A. Pollack and Michael L. Wachter

current employment. Women who have raised children are expected to accumulate less total work experience and therefore less market-oriented human capital than men or childless women.[16]

It is also straightforward to use the analytical framework sketched above to show that an exogenous increase in the market wage rate, perhaps associated with an increase in the capital intensity of industrial production, may induce a shift into more goods-intensive household production as well as more goods-intensive consumption, and an increase in the supply of labor services to the market on the part of household members who had previously been relatively specialized in household production. At the same time, total household demand for leisure in consumption may increase in response to increases in full income associated with higher wage earnings.

Motivated in part by this analytical framework, a relatively small but influential group of applied researchers has examined the relationship between households' commodity demand and labor supply decisions.[17] In most cases of which I am aware the relationship

---

(1975), "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," Journal of Political Economy, 83(2), pp. 255-278.

[16] It is interesting to note that Lundberg estimates simultaneous systems of labor supply equations for husbands and wives, and finds that the cross-effect of hours supplied by the spouse are significant only among households with young children. See Shelly Lundberg (1988), "Labor Supply of Husbands and Wives: A Simultaneous Equations Approach," Review of Economics and Statistics, 70 (2) pp. 224-235.

[17] M. Abbott and O. Ashenfelter (1976), "Labour Supply, Commodity Demand and the Allocation of Time," Review of Economic Studies, 43, pp. 389-411; William Barnett (1979), "The Joint Allocation of Leisure and Goods Expenditure," Econometrica, 47(3), pp. 539-563; Richard Blundell and Ian Walker (1982), "Modelling the Joint Determination of Household Labour Supplies and Commodity Demands," The Economic Journal, 92, pp. 351-364; Angus Deaton and John Muellbauer (1981), "Functional Forms for Labor Supply and Commodity Demands with and without Quantity Restrictions," Econometrica, 49 (6), pp. 1521-1532; W.E. Diewert (1974), "Intertemporal Consumer Theory and the Demand for Durables," Econometrica, 42(3), pp. 497-516; Louis Phlips (1978), "The Demand for Leisure and Money," Econometrica, 46(5), pp. 1025-1043. I am indebted to Erwin Diewert for recommending the paper by Barnett.

between the demand for durable goods and households' supply of labor services has been found to be significant. These results suggest that stocks of durable goods may be explanatory variables in labor supply decisions, particularly in the case of working-aged women. Other noteworthy contributions in this area include a recent theoretical essay by MaCurdy, as well as recent macroeconomic analysis that incorporates household production into simple long-run growth models.[18]

Beyond the general appeal of the general concept for the quantitative analysis of productivity growth, the literature on household production has influenced this research project in two specific ways. First, among the probit equations that are used to construct selection-bias correction factors with which the hedonic wage equations are estimated, one experimental specification includes the estimated probability of homeownership as an explanatory variable. The SIPP does not collect detailed information on the purchase of durable goods, but it does ask whether the household residence is owned by a family member. Recognizing that Consumer Expenditure Survey data show higher levels of durable goods expenditures among homeowners, in comparison with renters, the estimated probability of homeownership is interpreted here as an "instrument" for the

[18] Thomas E. MaCurdy (1999), "An essay on the Life Cycle: Characterizing intertemporal behavior with uncertainty, human capital, taxes, durables, imperfect capital markets, and non-separable preferences," Research in Economics, 53, pp. 5-46; Jess Benhabib, Richard Rogerson and Randall Wright (1991), "Homework in Macroeconomics: Household Production and Aggregate Fluctuations," Journal of Political Economy, 99, pp. 1166-1187; Ellen R. McGrattan (1998), "A Defense of AK Growth Models," Federal Reserve Bank of Minneapolis Quarterly Review, 22 (4), pp. 13-27; McGrattan, Rogerson and Wright (1997), "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," International Economic Review, 38, pp. 267-290.

capital intensity of household production that allows for the likely simultaneity of the homeownership and labor supply decisions.[19]

The second way in which the household production framework influences the estimating equations discussed below is through the inclusion of family income from financial assets, normalized by the poverty cutoff to offset variations in the demand for liquidity associated with variations in family size, among the explanatory variables in the experience equation. Normalized income from financial assets incorporates at least three effects: a wealth effect, expected to be negatively related to total work experience, a liquidity effect that might be negatively related to accumulated work experience of wives but positively related to the accumulated work experience of household heads, and a positive life cycle effect since individuals approaching retirement often shift the composition of their assets in favor of greater liquidity. It is often argued that wealth is endogenously determined over the course of the life cycle, and tests for the exogeneity of this variable are indicated. But the maintained hypothesis in the work reported here is that the flow of income from financial assets is exogenously determined by financial market conditions.

[19] As Matsayuma has noted, the domestic housing stock is one of the most important state variables in the economy; see Kiminori Matsuyama (1990), "Residential Investment and the Current Account," Journal of International Economics, 28, pp. 137-153. For an interesting recent discussion of this issue, see Jesus Fernandez-Villaverde and Dirk Krueger (2001), "Consumption and Saving Over the Life Cycle: How Important are Consumer Durables?" photocopy. As then latter authors note, homeownership may also capture a wealth effect since an owned home is the primary asset held by a large share of U.S. households. The incorporation of the homeownership probit into the systems of equations discussed here is still quite rudimentary. But the fact that homeownership may be a predetermined variable for estimation purposes makes it an attractive variable for applications such as this that focus on long-run relationships. Exogeneity tests and a careful specification of the stochastic relationship among the four equations of this prototype system appear to be indicated for the next stage of this research project.

General Structure of Equations Estimated

The general approach to model specification that is pursued in this paper is reminiscent of Blinder's seminal work on wage discrimination. Blinder compares reduced form and structural wage equation coefficient estimates.[20] His structural model can be specified in the familiar general form $Y = YB + XC + U$, where Y is a vector of endogenous variables, X is a vector of exogenous variables that contains information on family background as well as current exogenous variables, and U is a multivariate normal. The variables that Blinder includes in Y are the current wage rate, w, and categorical variables for education level, occupation, vocational training, union membership, veteran's status and tenure on the present job. Identification is achieved by imposing the assumption that the overall system has a block recursive structure. That is, the current wage rate is assumed to be determined by the other endogenous variables of the system, it is assumed that w is not a determinant of the other dependent variables in Y, and the error term of the wage equation is assumed to be stochastically independent of stochastic terms of the other six equations.

The words "predetermined variables" do not appear in Blinder's article, and the variables do not have time subscripts. Nonetheless the temporal sequence in which decisions regarding the endogenous variables are usually made clearly motivates his structural specification.[21] In this respect it is comparable Heckman's early work on heterogeneity

[20] Alan S. Blinder, "Wage Discrimination: Reduced Form and Structural Estimates," cited above.
[21] For example, Blinder writes "In the intuitive model I have in mind, each individual is presented with endowments of human and non-human capitals and at some point in the life-cycle, jointly determines how far he wishes to pursue his formal education and to what occupational strata he aspires." Blinder (), "Wage

and state dependence, in which retrospective information on pre-survey work experience and work experience reported during the course of successive interviews are treated asymmetrically.[22] In subsequent work focused on the estimation of hours equations, Mroz tested the null that marital status, number of children, and work experience are exogenous. He did not reject the null in the case of relatively unrestrictive "generalized Tobit" estimation procedures that appear to be comparable to the estimation procedures examined here. Thus there is influential precedent for this approach.

As was the case for the structural wage equation examined by Blinder, the wage equations estimated for this research project can be organized within the context of a block-recursive structure in which different equations reflect optimization of the endogenous variables over different, finite time horizons.[23] The predetermined variables in this specification are work experience at $t-5$, marital status, number of children, schooling level, and accumulated financial assets.[24] Current total work experience, the probability of employment in the current period, and the current wage equation are assumed to form a block of equations that is stochastically independent of predetermined

---

Discrimination," cited above, p. 441. Also see the early work of Chamberlin and Griliches, in which family background variables were found to be significant determinants of school achievement, but less important determinants of earnings once school achievement is taken into account.

[22] This distinction between pre-survey work experience and employment spells reported during the course of the survey is not acknowledged by Mroz, who criticizes this asymmetry. See Heckman () "Heterogeneity and State Dependence", TK, and T. A. Mroz (1987), "The Sensitivity of and Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica, 55, pp. 765-799. I am indebted to Margorie Honig for recommending that I examine Mroz's influential article.

[23] This general approach to the identification of structural systems of equations is illustrated, in an application involving the estimation of dynamic systems of factor demand equations with longitudinal microdata from the LRD, in Moeller (1995), Systems of Factor Demand Equations Derived from a Model of Monopolistic Competition: Results from Time Series Cross Section Data. Doctoral Dissertation for the Columbia University Economics Department, Ann Arbor: University Microfilms.

[24] Predicting years of school completed, which are generally assumed to be a function of parents' income and schooling levels, is beyond the scope of the current work. Therefore schooling is treated as a

and family background variables. In addition, this three-equation system is expanded to include a fourth equation that represents the probability of living in a home owned by a household member, as noted above. Incorporation of the homeownership probit is discussed in an earlier paper and not revisited here, but the "structural" specification of the selection bias correction factor that is incorporated in the wage equations reviewed below.[25]

To fix ideas, these systems of equations may be specified in general terms as follows.

$$\varphi_{EX}(EX_t) = G_{EX}[School, FAM, Health, Wealth, Nowrkspl, DEMOG, \varepsilon_E]$$

$$\Pr(EMP) = G_{EMP}[School, EX_{t-5}, FAM, Health, Wealth, DEMOG, GEO, \varepsilon_H].$$

$$\varphi_W(W_t) = G_w[School, \varphi_{EX}(EX_t), PT, Sector, Blue, Race, \lambda, \varepsilon_w]$$

In some cases $\varphi(Y) = Y$, and in other cases $\varphi(Y) = \ln Y$. $School$ is either a set of categorical variables defined in terms of years of school completed, or a linear spline in years of school completed with spline knots at conventional degree-completion years. $FAM = (C, W_{SP})$ contains information about the respondent's family that is expected to affect his or her labor supply decision; the variables denoted C reflect the number of children of the respondent and/or the respondent's spouse (in some cases the number of young children), and a variable W_{SP} that takes the value of the wage rate of the spouse if one is present and employed, and zero otherwise. $Health$ takes a value of 1 if the

predetermined variable, and the hypothesis that additive unobserved individual effects such as ability in school are uncorrelated with unobserved ability and thus accumulated work experience is maintained.

[25] See Moeller (1999) for details. At this stage this experimental specification is not very refined; the term "structural" refers more to the intent of the specification than to its current implementation, although the specification is formally identified.

respondent reports a disability that interferes with his or her ability to hold a job, and zero otherwise. *Wealth* is calculated as family income from financial assets divided by the official poverty cutoff; this normalization is intended to result in a variable that represents the transactions demand for liquidity, as noted above, and reduces the degree of heteroskedasticity that might otherwise be introduced by the financial income variable. *Nonwrkspl* is a vector that includes a categorical variable that takes a value of 1 if the last spell of non-employment for 6 months or longer was involuntary unemployment, and a variable that is equal to the duration of the last spell out of employment if it was involuntary, and zero otherwise.

DEMOG is a subvector that contains the following demographic variables: *Age* is the respondent's age minus 65 if that difference is positive, and zero otherwise. *Ms* is a categorical variable that reflect the respondent's marital status. *RET* is a linear spline function in age with a single knot at 55, intended to capture possible approach of retirement and the associated depreciation of market-oriented human capital. *GEO* is a subvector of geographic variables that includes categoricals for Census regions, and a variable that takes a value of 1 if the respondent lives in a city and zero otherwise.

PT takes a value of 1 when the respondent usually works less than 35 hours a week, and 0 otherwise. *Blue* takes a value of 1 when the respondent is a male production worker, and 0 otherwise. *Sector* is a vector of categorical variables that identify the major industrial sector in which the largest share of total hours was worked in a given calendar year. The selection bias correction factor λ is an inverse Mills ratio. The vector of stochastic terms, $\varepsilon = \left(\varepsilon_{EX_t}, \varepsilon_{EMP_t}, \varepsilon_{w_t}\right)$ is assumed to have a mean zero, and it is assumed that $\sigma_{EX_{t-5}, EMP_{it}} = 0$. Exclusion restrictions that serve to identify this system, subject to the normalization required for the selection bias factor, are outlined in Appendix B.

Specification and Estimation of Experience Equation.

The explanatory variables that enter the experience equation that is used in the current official index consist primarily of discrete counterparts of variables that also enter the wage equation. Separate identification of the coefficients of the experience and wage equations depends on the fact that the variables in the experience equation consist of mappings from two continuous variables (years of schooling and number of children in the family) to sets of discrete categorical variables that are defined in terms of the same two variables, a potential experience variable that is itself a linear combination of age and years of school completed, and "piecewise" variables obtained by multiplying the categorical variables by potential experience. This approach is not ideal for the purposes of structural or "causal" analysis. Consequently several alternative specifications of the work equation have been examined, in all cases with an eye toward the construction proxies or "instruments" that might also be used in two-sample estimation procedures.

The alternative specifications estimated to date include a linear spline in years of school completed, a demographically-driven quadratic that follows an early specification of Heckman, a specification in which the log of actual work experience is the dependent variable following work by Lancaster and Chesher, a more short-run specification that incorporates lagged work experience as a predetermined explanatory variable, and an S-curve that has been used in the time series literature to study market saturation. Detailed descriptions of the equations estimated are included in Appendix D.

Although the extensive literature on the estimation of duration may not be directly applicable to the problem at hand, for reasons discussed below, but it is not entirely clear that it is inapplicable either. For this reason it may be worth noting that although semi-parametric specifications may be preferred for the purposes of estimating the duration of completed spells, there does not appear to be a consensus in the literature on the criteria for preferring one semi-parametric approach to estimation over another. Furthermore the entire productivity series relies on separability assumptions that embody well-known functional form restrictions, and in this sense continued reliance on functional form restrictions is internally consistent.[26] On both of these counts the specification of a Weibull distribution for the estimation of systems of wage and employment duration equations as suggested in the work of Lancaster, and Lancaster and Chesher is particularly attractive.[27]

---

[26] Ernst R. Berndt and Christensen (1973), "The Internal Structure of Functional Relations: Separability, Substitution and Aggregation," Review of Economic Studies, 40 (3), pp. 403-410.

[27] T. Lancaster and A.D. Chesher (1984), "Simultaneous Equations with Endogenous Hazards," in Studies in Labor Market Dynamics, G.R. Neumann and N. Westergaard-Nielsen, Eds, Berlin: Springer-Verlag; Lancaster (1985), "Simultaneous Equations Models in Applied Search Theory," Journal of Econometrics, 28, pp. 155-169, Lancaster (1986), "Some Remarks on Wage and Duration Econometrics," in Unemployment, Search, and Labor Supply, Blundell and Walker, Eds., Cambridge: Cambridge University Press, and Lancaster (1990), The Econometric Analysis of Transition Data, Econometric Society Monograph No. 17, Cambridge: Cambridge University Press.

In this approach specifications experience and wage equations take the following form:

$$E(t) = \phi_1(t)\exp^{\left(X_1(t)\beta_1 + \varepsilon_1\right)}$$

$$W(t) = \phi_2(t)\exp^{\left[\left(X_2(t), E(t)\right)'\beta_2 + \varepsilon_2\right]}$$

The variable t represents potential experience, or years since completion of school. In keeping with the general rationale discussed above, it is assumed that the variables in the vector $X_1(t)$ are time-varying but exogenous, in the sense defined by Lancaster in his monograph on transition data.[28] Taking logarithms, as is standard practice in the estimation of human capital wage equations, is intuitively appealing in view of the well-recognized skewed distributions of wages and accumulated work experience. The result is a triangular system of the following form, where X_1 and X_2 may contain some variables in common:

$$\ln[E(t)] = \ln[\phi_1(t)] + X_1(t)\beta_1 + \varepsilon_1$$

$$\ln[W(t)] = \ln[\phi_2(t)] + [X_2(t), \ln E(t)]'\beta_2 + \varepsilon_2$$

This system is estimated most efficiently by maximum likelihood procedures, but implementation of maximum likelihood procedures in the context of selection-bias correction is not always straightforward, and least-squares should also yield consistent parameter estimates.[29] Since the logarithmic transformation is monotonic and the distribution of $\ln[E(t)]$ is roughly comparable in shape to the distribution of $\ln[W(t)]$,

[28] See Lancaster (1990), page 28.

[29] Lancaster (1983), "Generalized Residuals and Heterogeneous Duration Models: The Exponential Case," Bulletin of Economic Research, 35 (2), pp. 71-85; Jeffrey M. Woodridge (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge MA: MIT Press, p. 566.

entering the predicted value of the log of accumulated experience into the second estimating equation should help decrease residual heteroskedasticity. It is interesting to note that the R^2s of the equations with the log of potential experience on the left-hand-side are substantially lower than their counterparts with experience measured in levels. But the performance of the predicted log experience variable in the wage equation is among the strongest of the specifications examined.

An alternative approach to equation specification, also concerned with possible correlations in the residuals of the experience and wage equations, involves the use of a lagged value of actual accumulated work experience as an explanatory variable in the experience equation. Since lagged work experience is a predetermined value it should be uncorrelated with current wage rates when contemporaneous shocks or short-lived random "individual effects" are the source of correlation between wage and experience equation residuals.[30] Therefore work experience accumulated by year $t-5$ has been entered in one of the alternative experience equation specifications. Endogeneity tests for the lagged work experience variable are clearly indicated. In prior work, Heckman found lagged work experience to be an endogenous explanatory variable in an experience equation, but exogenous in the wage equation. And recent research by Mincer and Danninger, using the PSID, suggests that technological shocks cause unemployment and employment rates to deviate from long-run equilibrium for periods of 3-5 years.[31] Thus it seems plausible that a five-year lag will be sufficient to justify treating the lagged

[30] Heckman (1981), "Heterogeneity and State Dependence," Studies in Labor Markets, Sherwin Rosen, Ed., Chicago: University of Chicago Press, pp. 91-139.
[31] Jacob Mincer and Stephan Danninger (2000), "Technology, Unemployment, and Inflation," NBER Working Paper No. 7818, http://www.nber.org/papers/w7817.

variable as a predetermined variable. It is noteworthy that although lagged actual work experience variable substantially increases the explanatory power of the experience equation relative to potential experience, the resulting predicted value for potential experience does not increase the explanatory power of the wage equation nearly as dramatically.[32]

An "S-curve," taken from the time series literature on market saturation, was also estimated. In this case the inverse of potential experience was entered on the right-hand-side, without a quadratic term. In future work a quadratic might usefully be considered: when it is omitted predicted work experience values are quite high at the upper tail. Nonetheless, the within-cell wage equation prediction error associated with this specification is often fairly small.

Unobserved, randomly-distributed but sustained individual effects such as ability may also result in correlation among the stochastic components of the schooling, work experience and wage variables, as recognized at least since the 1970's.[33] It has been argued that persons with more innate ability will naturally achieve greater success in school and in the workplace than their less able colleagues, and that OLS estimates of the returns to schooling and work experience will be biased as a result.[34] The use of predicted actual work experience rather than reported values as explanatory variables in the wage equations is motivated in part by this concern, and in part by an institutional interest in extending this research to applications involving two-sample estimates. It is

[32] Neuman reports similar results in TK.
[33] Griliches and Chamberlin.

noteworthy in this regard that Krueger and Angrist reject the null hypothesis of significant unobserved ability in the wage equation, in their work with split samples.[35] Nonetheless the systematic implementation of exogeneity tests within the context studied here may be indicated for the next stage of this project.

The work history topical module of the SIPP collects information on both single and multiple employment spells, and these spells may be completed or censored . Observations on wage earners, who were employed during the period spanned by the survey, are censored in the sense that the term is used in the econometric literature on the estimation of duration. Especially among women, spells of employment were multiple spells. On average across all panels roughly 60% of male respondents reported having been employed at least 6 months per year every year since the first year of employment lasting 6 months or longer; the corresponding percentage for females is roughly 30%. Therefore, by necessity, for the purposes of this study persons who report having 6 or more months of continuous employment per year are considered to have been employed continuously. Observations with at least one spell of non-employment lasting 6 months or longer are observations with multiple employment spells. Among respondents aged 21-64 with at least one spell of non-employment that lasted 6 months or longer, information is collected on the duration of the most recent spell, and on the total number of spells.

---

[34] This concern motivates the work by Angrist and Krueger (1992, 1995), cited above.

[35] Joshua Angrist and Alan Krueger (1995), "Split Sample Instrumental Variable Estimates of the Returns to Schooling," Journal of Business and Economic Statistics, Vol. 13, No. 2, pp. 225-235. Also see

As noted above, it is not obvious that the extensive literature on duration estimation should be brought to bear in this research, since the variable of interest is the total number of years worked to date, rather than the expected duration of all employment spells. This is because, within the analytical framework of standard static human capital wage equations, it is assumed that employees accumulate both general and specific human capital throughout their working lives. It is this "stock" of accumulated human capital that is assumed to increase current worker productivity. Thus the censored values are the variables of interest for the purposes of estimating human capital wage equations, especially since wage data are unavailable by definition for those whose lifetime employment has been completed. Similarly, while employment during one's lifetime could be specified within a censored regression framework, a selection bias correction factor has not been incorporated in the experience equation because only about 6% of females and less than 1% of the males in the population report that they never worked.

Regression and Labor Composition Index Number Results

Annual coefficient estimates and associated summary statistics for each of the experience equations estimated are reported separately by gender in Appendix D. Two basic wage equation specifications are estimated: the current BLS specification described in Section III, and a new specification is the revised specification discussed in Section V. The current BLS specification includes information on family structure and regional categorical variables in the wage equation while the alternative specification does not.

Griliches (1970), "Notes," cited above, and Blinder (TK), "Wage Discrimination: Reduced Form and Structural Estimates," The alternative specification includes broadly-defined categorical variables for primary industry of employment and a selection bias correction factor, neither of which is incorporated in the current BLS specification. In the case of the alternative specification, information on household composition enters the wage equation indirectly, through the selection bias correction factor.

More specifically, predicted values for actual work experience are obtained with each of five experience equation specifications, as discussed above. The wage equation used to construct the current index is estimated with predicted experience values based on the experience equation currently used to construct the index, and with potential experience. In addition, each of the four alternative specifications of the experience variable discussed above is entered as a proxy in the alternative wage equation. Two alternative selection bias correction factors, denoted λ, are included with each of the new alternative specifications of the wage equation/experience proxy. The first selection bias correction factor, denoted λ_{BA}, is estimated with the coefficients of a simple "baseline" probit that follows Heckman's early work. The second selection bias correction factor, denoted λ_{ST}, is estimated with the coefficients of a probit in which the predicted probability of homeownership replaces normalized income from financial assets as the wealth variable, the number of children under six years of age and the number of school-aged children are entered as separate explanatory variables, and categorical variables for nine Census regions are included to capture differences in local labor market conditions.

Examination of the tables in Appendix D reveals that the current BLS experience equation specification is often dominated by the alternatives estimated, when evaluated in terms of the R^2s associated with the experience and wage rate equations. In addition, simple counts of the incidence with which predicted sample values fall above or below the range $(0,75]$ suggest that the current BLS specification does not fit the data as well in the outer tails of the distribution as several of the other specifications considered. (Tables and charts on outliers of predicted values are included in Appendix D, in a separate section between the experience equation and the wage equation coefficients.)

However, the wage equation parameter estimates are used to generate predicted wage rate values with which to construct the labor cost share weights $\omega_{l,t}$ for persons with characteristics l at time t, as discussed in Section III. The within-cell fit of the equations is of substantial interest for this reason. Summary statistics on weighted ratios of the antilog of the wage equation residual divided by the actual wage rate were therefore generated for cells defined in terms of alternative data partitions and 5-year age intervals.[36] That is, partitions defined in terms of broad ranges of actual work experience were divided into sub-cells that can be sorted by age, to facilitate direct comparison by cohort across partitions.

To provide a broad overview of the results, the weighted within-cell-and-cohort ratios were pooled across all years, and summary statistics were generated for each combination

[36] Bias-adjustment for the non-linear conversion from of the predicted log wage is not part of the current BLS procedure and has not been implemented in the results reported here. See Dhrymes (1995), "On Devising Unbiased Estimates for the Parameters of the Cobb-Douglas Production Function," reprinted in Theoretical and applied econometrics: The selected papers of Phoebus J. Dhrymes, Aldershott: Elgar.

of experience, participation, wage equation and data partition considered. Results are summarized in the tables below. The three rows that tend to have the lowest ratios are highlighted in each table.

The first highlighted row, denoted *optw*, corresponds to the partition used in the current series except that cells are defined over 5-year work experience intervals rather than 1-year intervals of potential experience. [37] The second highlighted row, denoted *newn*, corresponds to the alternative partition proposed in this paper except that data are partitioned by broad age intervals rather than by actual work experience.[38] The third highlighted row, denoted *neww*, corresponds to another alternative partition that also sorts workers into broad intervals of actual work experience. Thus the first and third highlighted rows identify partitions that are only feasible when the index is constructed by partitioning the data in terms of the actual work experience information in the SIPP, while the second highlighted row denoted *newn* identifies a new alternative partition that is feasible with TS2SLS or TSIV estimation procedures.

[37] The current OPT index, denoted *optp* in the tables, is constructed by partitioning microdata from the March annual supplement of the CPS into 72 age vales (0-1, 1-2, ..., 71 and over), 7 discrete categorical variables for years of school completed (fewer than 4 years, 4-8, 8-11, 12, 12-14, 14-16, 16, over 16), and gender. For females, there are also 4 discrete categorical variables for number of children (none, 1, 2, more than 2), and marital status (ever married, never married). Potential experience, number of children, and the "ever married" variable jointly serve as an estimate of actual work experience. See BLS Bulletin 2426 (1993), cited above, Table D-1, p. 68. The partition denoted *optw* is the same, except that 5-year intervals of actual work experience replace the 1-year potential experience intervals. The partition denoted *opta* is the same except that 5-year age intervals replace the 1-year potential experience intervals.

[38] For the partition denoted *new*, cells are defined in terms of 4 broad regions (northeast, south, central and west); 3 broad industry categories (manufacturing, mining, transportation and agriculture; finance and wholesale and retail trade; and services), 2 broad age categories (younger than 55, and 55 and older) 3 broad schooling categories (less than 12 years, 12 years, 13-16 years, and more than 16 years of school completed), and gender. Among male employees in heavy industry there is a further partition into production and non-production employment that follows Berndt and Christensen (1974), cited above. The

In the case of males, abstracting from partitions in terms of actual work experience and focusing on the row denoted *newn* in which data are partitioned by industry, schooling, and broad age intervals, the current experience and wage equation without a selection bias correction result in the smallest mean and median ratios. The new wage equation specification, with the predicted level of actual experience on the RHS, the "structural" selection bias correction factor and the same data partition yields the smallest ratio at the 25^{th} percentile. The inter-quartile range and 75^{th} percentile ratios are smallest when potential experience replaces estimated experience in the current wage equation and the selection bias correction factor is omitted.

For females, again abstracting from partitions in terms of actual work experience, the smallest mean within-cell-and-cohort ratio, and the smallest value at the 25^{th} percentile, are associated with the partition *newn* together with a wage equation in which potential experience replaces predicted actual experience and there is no selection bias correction. The same partition by industry, schooling and broad age intervals in combination with the Weibull specification for the experience equation and the structural specification of the participation equation results in the smallest median ratio. The same partition in combination with the current BLS experience equation and no selection bias correction performs takes the lowest value at the 75^{th} percentile, and has the smallest interquartile range

partition denoted *newn* differentiates between employees with undergraduate and advanced degrees. The partition denoted *neww* incorporates a further partition into 3 broad intervals of actual work experience.

FEMALES

EQN	1 OPT/PE	2 OPT/opt	3 WB/BA	4 WB/ST	5 HC/BA	6 HC/ST	7 LV/BA	8 LV/ST	9 LG/BA	10 LG/ST	11 SC/BA	12 SC/ST
	MEAN WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE											
PARTN												
optp	0.15392	0.15447	0.15421	0.15418	0.15301	0.15289	0.15373	0.15372	0.15274	0.1526	0.15368	0.15369
optw	0.14964	0.14981	0.1503	0.15005	0.14901	0.14865	0.14992	0.14968	0.15082	0.15067	0.15037	0.1502
opta	0.16392	0.16492	0.16611	0.16611	0.16212	0.16186	0.16546	0.16549	0.16407	0.16389	0.16561	0.1657
new	0.1561	0.15668	0.15614	0.15618	0.15674	0.1569	0.15637	0.15642	0.1558	0.15592	0.15551	0.15557
newn	0.14672	0.14684	0.14697	0.14699	0.14742	0.14754	0.14702	0.14703	0.14801	0.14811	0.14678	0.14684
neww	0.13802	0.13775	0.13935	0.13914	0.13924	0.13909	0.13947	0.13926	0.1416	0.14178	0.13957	0.13946
	75TH PERCENTILE WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE											
optp	0.18402	0.18486	0.18279	0.18271	0.18266	0.18272	0.18247	0.1822	0.18019	0.18014	0.18201	0.1819
optw	0.17973	0.17929	0.17733	0.17715	0.17763	0.17738	0.17672	0.1767	0.17714	0.17716	0.17735	0.17737
opta	0.19787	0.19864	0.19873	0.19854	0.19576	0.1957	0.19715	0.19744	0.19602	0.1956	0.19803	0.19807
new	0.18468	0.18492	0.18588	0.18572	0.187	0.18707	0.18608	0.18607	0.18466	0.18453	0.18456	0.18482
newn	0.17622	0.17549	0.17679	0.1766	0.17715	0.17734	0.17637	0.17666	0.1781	0.17787	0.17632	0.1768
neww	0.16604	0.16494	0.16639	0.16599	0.16688	0.16667	0.16629	0.16586	0.16922	0.16917	0.16707	0.16694
	MEDIAN WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE											
optp	0.14537	0.14499	0.14442	0.1444	0.1436	0.1435	0.14383	0.14352	0.14362	0.14342	0.14435	0.14442
optw	0.1415	0.14066	0.14088	0.1404	0.1404	0.1401	0.14031	0.13982	0.14239	0.14212	0.14153	0.14111
opta	0.15543	0.15517	0.15418	0.154	0.1537	0.1534	0.1532	0.15297	0.15166	0.1516	0.15369	0.15354
new	0.15022	0.14944	0.14952	0.1495	0.1485	0.1486	0.14962	0.14979	0.14982	0.14987	0.14915	0.14941
newn	0.14221	0.14178	0.14175	0.1415	0.1416	0.1417	0.1416	0.14176	0.14392	0.14389	0.14203	0.14208
neww	0.13232	0.1317	0.1334	0.1331	0.1328	0.1325	0.13363	0.13312	0.13661	0.13668	0.13412	0.13407
	25TH PERCENTILE WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE											
optp	0.11238	0.11262	0.11317	0.1131	0.11179	0.11169	0.11275	0.1125	0.1128	0.1128	0.11308	0.11295
optw	0.10909	0.10942	0.1105	0.11017	0.10927	0.10896	0.11006	0.1098	0.1116	0.1114	0.11075	0.11045
opta	0.11883	0.1191	0.11888	0.11887	0.11748	0.11741	0.11864	0.1182	0.1183	0.1182	0.11868	0.11877
new	0.11911	0.11878	0.11846	0.11861	0.11688	0.11708	0.1186	0.1188	0.1199	0.1198	0.11883	0.11896
newn	0.11063	0.11095	0.1116	0.11175	0.11071	0.11077	0.11168	0.1117	0.1128	0.113	0.11096	0.11113
neww	0.10253	0.10283	0.10437	0.10437	0.1034	0.10341	0.1047	0.1047	0.1057	0.1061	0.10421	0.10443
	INTER-QUARTILE RANGE: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE											
optp	0.07165	0.07224	0.06962	0.06961	0.07087	0.07103	0.06972	0.06963	0.06736	0.06731	0.06894	0.06895
optw	0.07063	0.06987	0.06682	0.06698	0.06836	0.06842	0.06665	0.06689	0.06551	0.06573	0.0666	0.06693
opta	0.07904	0.07954	0.07985	0.07967	0.07827	0.07829	0.0785	0.07922	0.07767	0.07737	0.07935	0.0793
new	0.06557	0.06614	0.06742	0.06711	0.07012	0.06999	0.06748	0.06724	0.06467	0.06469	0.06573	0.06586
newn	0.06558	0.06454	0.06519	0.06485	0.06643	0.06657	0.06469	0.06488	0.06527	0.06486	0.06536	0.06567
neww	0.06351	0.06211	0.06202	0.06162	0.06348	0.06326	0.06159	0.06108	0.06349	0.06306	0.06286	0.06251

MALES

	1	2	3	4	5	6	7	8	9	10	11	12
EQN	OPT/PE	OPT/opt	WB/BA	WB/ST	HC/BA	HC/ST	LV/BA	LV/ST	LG/BA	LG/ST	SC/BA	SC/ST

MEAN WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE

PARTN	1	2	3	4	5	6	7	8	9	10	11	12
optp	0.11167	0.11203	0.11043	0.1103	0.10866	0.10835	0.11015	0.10998	0.10898	0.10899	0.10972	0.10954
optw	0.10789	0.10822	0.10715	0.10682	0.10476	0.10424	0.10675	0.10641	0.10773	0.10765	0.10713	0.1068
opta	0.12128	0.122	0.11938	0.11909	0.11543	0.11494	0.11897	0.11869	0.1172	0.11713	0.11876	0.11845
new	0.10717	0.10738	0.10972	0.1097	0.10976	0.10977	0.10989	0.10989	0.10937	0.10949	0.10965	0.10955
newn	0.09975	0.09972	0.10233	0.1023	0.10181	0.10181	0.10226	0.10224	0.10435	0.10441	0.10288	0.10277
neww	0.09427	0.09417	0.09651	0.09645	0.09547	0.09535	0.09638	0.0963	0.09843	0.09856	0.09699	0.09691

75TH PERCENTILE WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE

PARTN	1	2	3	4	5	6	7	8	9	10	11	12
optp	0.13276	0.13352	0.13085	0.13056	0.12938	0.12903	0.1305	0.13036	0.12579	0.12559	0.12906	0.1287
optw	0.12869	0.12881	0.12678	0.12649	0.12485	0.1244	0.12623	0.1261	0.12474	0.12485	0.12642	0.12602
opta	0.14652	0.1477	0.14291	0.14283	0.13754	0.13694	0.14264	0.1422	0.13762	0.13756	0.14173	0.14153
new	0.12596	0.12627	0.12966	0.12986	0.13099	0.13099	0.12983	0.12981	0.12689	0.127	0.12887	0.12891
newn	0.1155	0.11563	0.12072	0.1207	0.12089	0.12076	0.1205	0.12082	0.1222	0.1221	0.12165	0.1215
neww	0.10955	0.10898	0.11313	0.11309	0.11295	0.11299	0.11273	0.11269	0.11551	0.11549	0.11413	0.11416

MEDIAN WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE

PARTN	1	2	3	4	5	6	7	8	9	10	11	12
optp	0.10433	0.10371	0.10158	0.10143	0.1017	0.10161	0.10108	0.10101	0.10107	0.10107	0.10187	0.10176
optw	0.10063	0.0999	0.09803	0.09766	0.09824	0.09772	0.09751	0.09701	0.10003	0.09997	0.09939	0.0989
opta	0.11413	0.11341	0.10954	0.10896	0.10922	0.10864	0.10868	0.10868	0.10739	0.10716	0.10964	0.10917
new	0.09962	0.09935	0.10201	0.1022	0.10179	0.10202	0.10208	0.10208	0.10231	0.10246	0.10298	0.10301
newn	0.09344	0.0933	0.09586	0.09585	0.09579	0.09576	0.0957	0.09578	0.09827	0.09825	0.09692	0.09699
neww	0.08789	0.0874	0.08952	0.08937	0.08893	0.08865	0.0895	0.08945	0.09239	0.09232	0.09089	0.09071

25TH PERCENTILE WITHIN-CELL RATIO: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE

PARTN	1	2	3	4	5	6	7	8	9	10	11	12
optp	0.08076	0.08081	0.07988	0.07979	0.07892	0.07876	0.07982	0.07968	0.08118	0.08106	0.08047	0.08036
optw	0.07811	0.07814	0.07788	0.07693	0.07625	0.07581	0.07692	0.07693	0.08005	0.07983	0.0778	0.07752
opta	0.08622	0.08637	0.08438	0.08382	0.08263	0.08254	0.08373	0.08358	0.08528	0.08516	0.08504	0.08478
new	0.08104	0.08067	0.08155	0.08121	0.08031	0.08264	0.08163	0.08152	0.08285	0.08289	0.08236	0.08207
newn	0.07544	0.07553	0.07548	0.07547	0.07548	0.07545	0.07586	0.07193	0.07807	0.07813	0.07599	0.07585
neww	0.07017	0.07024	0.07358	0.0703	0.06973	0.06972	0.07587	0.07045	0.07232	0.07242	0.07067	0.07066

INTER-QUARTILE RANGE: ANTILOG WAGE EQUATION RESIDUAL TO ACTUAL WAGE

PARTN	1	2	3	4	5	6	7	8	9	10	11	12
optp	0.052	0.0527	0.0597	0.05077	0.05046	0.05278	0.05046	0.05028	0.05068	0.05068	0.04461	0.04453
optw	0.05059	0.05067	0.04976	0.04956	0.0486	0.04859	0.0486	0.04859	0.04932	0.04917	0.0447	0.04502
opta	0.0603	0.06132	0.05853	0.05902	0.05491	0.0544	0.05491	0.0544	0.05891	0.05862	0.05234	0.0524
new	0.04492	0.0456	0.04812	0.04865	0.05068	0.0573	0.05068	0.05073	0.0482	0.04828	0.04403	0.04412
newn	0.04007	0.04008	0.04524	0.04522	0.04541	0.04531	0.04541	0.04531	0.04464	0.04489	0.04413	0.04397
neww	0.03938	0.03874	0.04277	0.04279	0.04321	0.04327	0.04321	0.04327	0.04214	0.04224	0.04318	0.04308

Mneumonics for Columns 1-12 in Above Tables
OPT/PE identifies the current wage equation specification, with potential experience replacing the actual work experience proxy. OPT/opt identifies the current approach, with coefficients from the contemporaneous experience equation used to construct a work experience proxy. In the remaining mneumonics, WB identifies an equation with the log of actual experience on the LHS and the log of potential experience on the RHS. HC follows an early specification in by Heckman with the level of actual experience on the LHS, LV is a new specification with levels on the LHS, LG identifies an experience equation with actual experience on the LHS and lagged experience on the RHS, and SC is an S-curve where the logs of actual and potential experience are entered on the LHS and RHS respectively. In columns 3-12, BA identifies selection bias correction factors based on an employment probit that follows the early work of Heckman. ST identifies selection bias correction factors based on a probit in which the probability of living in a home owned by a household members is entered as an index of the capital intensity of household production, and categorical variables for 9 Census regions capture regional differences in employment population ratios.

Partitioning in terms of broad schooling, industry and work experience intervals generally results in smaller within-cell-and-cohort ratios than the ones mentioned above. In the case of males, focusing on the partition labeled *neww*, the current wage equation specification with no selection bias correction yields the lowest mean, median and inter-quartile range, as well as the smallest ratio at the 75th percentile. In the case of females the current specification with no selection bias correction yields the lowest mean and median within-cell-and-cohort ratios. The wage equation specified in terms of potential experience, with no selection bias correction factor, yields the lowest ratio at the 25th percentile. The experience specification in levels, in combination with the alternative specification of the wage equation and the structural specification of the participation equation results in the smallest inter-quartile range.

These intermediate results do not substitute for conventional specification and exogeneity tests, which are clearly indicated if this research project is to be continued.[39] In addition, systematic review of these ratios by cohort may provide further insights into the strengths and weaknesses of alternative specifications. Nonetheless these preliminary results illustrate the importance of alternative data partitions to the quantitative results obtained with hedonic indexes.

Intuitively, however, consistency between the explanatory variables of the wage equation and the partitioning variables seems desirable in order to give analytical meaning to the separability assumptions that underlie the index number approach to productivity measurement. That is, it seems logical that the explanatory variables that are the true structural determinants of observed wage rates should also be the variables used to partition the data into groups within which relative wage rates are assumed to remain unchanged in the face of changes in prices paid for other factors of production.

Assuming that internal consistency between the explanatory variables of the wage equation and the variables employed in estimation is desired, the new specifications and partitions appear to dominate the current ones. This may be seen by comparing the 4 cells in the first two rows and first two columns of the upper left-hand portion of each panel, marked off as a box surrounded by a solid line, with the 20 cells in columns 3-12 and the last two rows, similarly boxed, in the bottom right-hand portions of each panel.

[39] Furthermore,it should be noted that tabulation of annual coefficient estimates revealed that some cross-sections variables intended to allow the slopes and interecepts of the wage equation to vary in the case of younger and older workers, from whom information on actual work experience is not collected, were omitted inadvertently. Although in most cases these variables are insignificantly different from zero when

The upper left-hand portions, i.e. the cells that correspond to the first two columns of the table and the rows denoted *optp* correspond to the wage equation and partition employed to construct the current BLS index. The row denoted *optw* is the same except that the partition is defined in terms of 5-year intervals of actual work experience rather than 1-year intervals of potential experience. The cells in columns 3-12 and the rows denoted *new*, *newn* and *neww* correspond to conditional mean cell wage rates, predicted with the 3- and 4-equation systems described above.

It is particularly noteworthy that the partition denoted *newn* does not rely on actual work experience data, but it is defined in terms of the key variables in the revised specification of the wage equation. It appears to dominate the current partition, and the revised partition defined in terms of 5-year age intervals. It is dominated in turn by the partition *neww*.

To recapitulate, key differences between the current and revised wage equation specifications are that the current approach does not incorporate a selection bias adjustment factor and it does not include categorical variables for industry of primary employment., while it does include regional categorical variables. The revised specification incorporates a selection bias adjustment factor. Regional categorical variables assumed to capture regional differences in local labor market conditions are entered in the employment probit that underlies the selection bias correction factor λ_{ST}, and categorical variables for industry of primary employment are entered in the wage

included, it is still true that the rankings of the ratios may well change after those errors have been corrected.

equation to capture industry differences in the physical- and specific-human-capital intensity of production that may result in substantial costs of adjustment.

Average index growth rates across all equation specifications examined are reported in the Table below, together with index growth rates calculated with cell mean growth rates. Growth rates for each partition/specification combination are reported in Appendix F, along with exact definitions of each partition.

Perhaps not surprisingly, given the different effects of the partitions considered on the ratios reported above, the choice of a data partition and the time period over which average annual growth rates are calculated both make a significant difference in the values taken by the labor composition index. The design of the 1984 work history topical module was substantially different from the design of subsequent topical modules, and no work history data were collected in the 1985 panel. Therefore average annual growth rates based on the period 1987-1993 are reported here.

For the period 1987-1993, following the current methodology, average annual index growth rates were 3 percent for males and 2.9 percent for females (cell OPTP, OPT in Appendix F). These values are quite similar to the index obtained with cell mean wage rates, which shows growth rates of 2.7 percent for males and females under the same data partition. In contrast, when data are broadly partitioned by schooling, industry and age (*newn*), average annual growth rates for males and females are much smaller (0.02 percent for males, and 0.08 percent for females on average across all equation specifications examined). When the labor composition index is constructed with cell mean wage rates rather than predicted values, the partition *newn* results in average annual growth rates of 0.04 percent for males and 0.10 percent for females.

Partition	Sex	Avg. Ann. Growth Hedonic	Avg. Ann. Growth Mean Wage	Partition	Avg. Ann. Growth Hedonic	Avg. Ann. Growth Mean Wage
optp	males	0.0309	0.0265			
	females	0.0294	0.0267			
opta	Males	0.0034	0.0031	*newn*	-0.0002	0.0004
	females	0.0026	0.0025		0.0008	0.0010
optw	Males	0.0441	0.0456	*neww*	0.0533	0.0539
	females	0.0680	0.0718		0.0466	0.0469

The effect of partitioning by actual work experience is striking. When data are broadly partitioned by schooling, industry and actual work experience (*neww*) the indexes average about 5 percent (5.33 percent for males and 4.66 percent for females on average, across all equation specifications). Alternative equation specifications again yield quite similar results. The index constructed with cell mean wage rates is 5.4 percent for males and 4.7 percent for females. This similarity between the cell mean growth rates and the hedonic index growth rates is reassuring because it suggests that the econometric estimates are not distorting the values that would be obtained under conventional index number procedures.

Conclusion

The results reported in this paper indicate that the ability to differentiate between workers with strong and weak labor force attachment is quantitatively important for the labor composition index, if a partition by accumulated total work experience is desired. Some analysts would argue that it is not appropriate to partition by variables that may be endogenous in the long run, such as primary industry of employment and total work experience, despite the fact that these partitions yield the smallest within-cell-and-cohort prediction errors.[40] However this paper has invoked the notion that lagged predetermined variables are "state variables," reflecting irreversible investments in physical and human capital, to motivate a recursive structural specification of experience, employment and wage equations. In the event that this recursive structure is not rejected by specification tests, then partitioning the data in terms of the state variables employed in estimation may be justified.

Alternatively, partitioning by predicted work experience might be more appropriate despite the errors in assigning observations to "bins" that would undoubtedly result. Partitioning by predicted work experience would have the advantage of supporting the two-sample estimation procedure currently envisioned by the BLS, while partitioning by actual work experience does not. Feedback from workshop participants on these points would be welcome.

As noted above, specification and exogenity tests are clearly indicated before evaluation of the alternative specifications discussed here will be complete. Nonetheless, and despite the fact that the choice of a particular set of equation specifications has a much smaller effect on the labor composition index than the choice of a partition does, the

[40] Glen Cain has stressed this point in email correspondence.

robustness of the equation parameter estimates in successive annual cross-sections speaks well to the usefulness of applying hedonic techniques to these data.[41]

Future work based on the household production framework might focus on estimating separate systems of wage and hours equations by household type in order to allow for greater simultaneity in the determination of wage rates and hours worked, and in the labor supply decisions of husbands and wives. It would be interesting to attempt to formulate separability tests based on flexible functional forms.[42] The estimation of separate systems of hours and wage equations for mature and elderly households, taking account of other income flows and asset holdings and more detailed account of health conditions, might also prove useful for an analysis of the effects of the labor supply decisions of the aging baby boom cohort on productivity growth.

This research has been aided substantially by the fact that the classical human capital wage equation is one of the most robust equations estimated in applied economics, and particularly well-suited to the application of hedonic techniques. Furthermore the data set examined was designed explicitly for the analysis of labor supply decisions. The application of hedonic techniques to markets undergoing rapid technological change, for which both theory and data may be less well-developed, is likely to be much more

[41] Some of this robustness is due to the fact that the annual cross-sections are estimated with pooled overlapping panels of the SIPP. For example, the estimates for 1987 are obtained by pooling data collected in the second year spanned by the 1986 SIPP panel and the first year spanned by the 1987 panel, where annual variables are averages or sums of monthly or quarterly reported values. A substantial amount of smoothing is undoubtedly achieved in the estimation process.

[42] Given the difficulty of identifying time spent in household production separately from time spent in consumption, application of the conventional approach would not be entirely straightforward. On this point see Pollack and Wachter, cited above. Prior conceptual work on the estimation of systems of equations with "mixed" data on prices and quantities might be useful in this context.

difficult. In such contexts, and particularly when the work is being undertaken by government statistical agencies accountable to the general public, the guidance provided by disinterested academic research can be invaluable.

Appendix A: Overview of Research Project and Background on the SIPP

Overall, the project has had the following objectives:

1. Examination of three sources of bias that may be present in the current measures. These sources of potential bias are: (a) systematic errors in the measurement of actual work experience, earnings, and hours worked, (b) sample selection bias, and (c) model misspecification bias.

2. Illustration of the usefulness of the econometric approach to the construction of index numbers, as it is applied in the construction of the labor composition index, through the development of research papers in which the structural determinants of changes in the labor composition index are analyzed in quantitative terms.

The SIPP is extremely well designed for the purposes of measuring changes in earnings and hours of employment at the national level. In addition to the availability of microdata on the total number of years worked by survey respondents, the following aspects of the SIPP's design are important for the construction of an index of shifts in the skill-composition of the work force.

First, the SIPP may provide improved estimates of quarterly and annual hours worked, relative to other large sets of microdata that are representative of the U.S. labor market.[1] In particular, it provides information on the number of hours worked by salaried employees that is not available from the Current Employment Statistics' (CES) establishment data.[2] The SIPP collects information on the number of weeks worked and the usual number of weekly hours worked, for each month in the calendar year, from all adult survey respondents. In contrast, the CES only collects information on hours worked by production and non-managerial workers during the pay period that includes the 12th day of the month. The SIPP also collects monthly information on the number of weeks worked and the number of hours worked per week on a second job.[3] Information on hours worked while self-employed in a first or second business is also collected from all adult SIPP respondents.

Second, the SIPP follows all members of a surveyed household for the duration of the 28-month survey period, even when some or all members of the household relocate. In contrast, the CPS does not follow individuals who separate from the surveyed household, and does not survey each household member for an entire calendar year. Thus short employment spells that might be omitted from retrospective information collected in the March CPS are likely to be incorporated in SIPP data. Similarly within-year selection bias of the type discussed by Hanoch should not be a problem, as it may be for estimates based on the annual March supplement to the CPS.[4]

---

[1] The strengths of SIPP estimates of total hours worked, relative to those available from the Current Population Survey (CPS) and Current Employment Statistics (CES) data, are discussed in Moeller (2002), cited above. This paper summarizes results from prior research in which known limitations of the hours estimates in the CPS are analyzed, as well as research that examines respondent recall error in PSID estimates of hours worked. The sequence of questions on hours worked in the PSID is quite comparable to the sequence of questions upon which the SIPP estimates are based.

[2] The BLS is currently working on ways to collect establishment data on hours worked by salaried employees. The resolution of this problem is not straightforward since many establishments do not maintain administrative records on hours worked by salaried employees.

[3] Data on hours worked on a second job are now collected in the CPS, but only from a fourth of the respondents, and only in selected months.

[4] Giora Hanoch (1980), "Hours and Weeks in the Theory of Labor Supply," in Female Labor Supply: Theory and Estimation, James P. Smith, Ed., Princeton: Princeton University Press, pp. 119-165. Also see Rebecca M. Blank (1988), "Simultaneously Modeling the Supply of Weeks and Hours of Work among Female Household Heads," Journal of Labor Economics, 6 (2), pp. 177-204.

Third, the precision of estimates of changes over time is known to be substantially increased with repeated observations on the same unit of analysis, since the relative importance of sampling variance is diminished with repeated observations. Consequently estimates of hours growth rates obtained with longitudinal data dominate estimates based on successive CPS cross-sections, all else equal, because repeated observations represent a smaller share of the total number of observations in the CPS.

Fourth, as noted above, the SIPP collects monthly information on weekly hours usually worked each month, number of weeks worked, and weekly earnings at two or more jobs, in interviews that are conducted every four months. This relatively short recall period is designed to minimize respondent recall error for a given sample size, relative to the retrospective questions in the annual March supplement of the CPS. It should also generate more precise estimates of cyclical fluctuations in total employment, hours worked and earnings than those available from the CPS March supplement, in which the primary frame of reference is the longest-held job during the previous calendar year.

Fifth, in addition to then-current (or then-recent) data collected at four-month intervals, retrospective information on each respondent's prior work history, migration, education and assets are collected in one-time "topical modules." These data support the estimation and analysis of structural models of household labor supply in which "gains from trade" may have been achieved through household members' specialization in market or non-market production.

It must be noted that the detailed and extensive information collected in the SIPP increases the amount of time required to assemble and process the data. Consequently the SIPP is much less timely than either the CES or the CPS, and probably cannot be used directly to construct quarterly indicators. However, the work described below demonstrates that it provides a very strong basis for the econometric analysis of hours and earnings growth rates, and it could prove useful for the construction of quarterly indicators within the foreseeable future. [5]

Initial Comparisons with Social Security Administrative Records.

The first stage of the project focused on a comparison of the SIPP's information on total accumulated work experience with parallel information in administrative records from the Social Security Administration (SSA) that had been matched to the SIPP microdata, in a prior research project undertaken by the SSA.[6] Administrative records on quarterly employment and earnings were compared directly with SIPP estimates of the total number of years worked. SIPP estimates were found to dominate the estimates based on administrative records due to incomplete SSA coverage during the early years of the program, and variability of program coverage rates over time.[7]

The work experience proxy currently used to construct the labor composition index, with the coefficients of an experience equation previously estimated with 1973 CPS microdata matched to SSA administrative records, has been shown to generate earnings profiles that are clearly steeper than the earnings profiles

---

[5] Prior to the introduction of computer-assisted interviewing procedures preliminary data were available roughly a year after the time period to which they refer. That is, data were available roughly 8 months after an interview, and the interview referred to the preceding 4 months. There is a greater current data-processing lag, but it is expected to decrease as the automated data collection procedures currently being developed increase in efficiency. Once the new data-processing procedures are working smoothly , preliminary data will probably be available a year or less from the month to which they refer.

[6] Some results from this prior project are reported in Howard M. Iams (1991), "Child Care Effects on Social Security Benefits," 1991 Annual Research Conference Proceedings, Bureau of the Census, pp. 255-271.

[7] The linking of administrative record and survey microdata also raises important privacy issues, and these have prevented the routine linking of comparable data sets in most other years. These privacy concerns are likely to preclude the reliable, routine linking of administrative record and survey microdata for some time to come.

generated with either the SSA or the SIPP 1984 work experience proxy. This result reflects the fact that the measurement error embodied in the SSA administrative records has changed over time, and/or it indicates that the structural relationship between actual work experience and wage rates has not been constant. Incorporation of a standard selection-bias correction factor into the 1984 wage equations estimated was found to lower female earnings profiles among younger married women, relative to earnings profiles from which the selection-bias correction factors were omitted. [8]

The initial rationale for the project was to use data from the 1984 panel of the SIPP, which had already been linked to SSA data, to test the null that the coefficients of the experience equation were stable over time. The statistical community has been interested in the potential cost savings associated with the increased use of administrative records, but remains concerned about confidentiality, comparability and data quality. [9] Since regular ongoing access to successive matches between survey data and administrative records is not assured, the stability of the experience equation coefficients over time was also an important question.

To evaluate the quality of each work experience measure, internal consistency checks were performed before the two estimates of respondents' total work experience were compared. In the case of the SSA data, information on the number of quarters of covered employment were compared with SSA data on quarterly earnings. In the case of the SIPP data, respondents' direct estimates of the total number of years worked were compared with estimates obtained by calculating the duration of each reported employment spell, and adding the number of years worked in each spell. From this internal standpoint, both data sets were found to be reasonably consistent.[10]

Comparison of the SIPP and the SSA measures of total work experience, in the table below, reveals that the SSA estimates for older workers were biased downward, in a manner consistent with relatively low Social Security coverage rates during the 1930's and 1940's. The SIPP estimates for workers younger than 21are understated because the full sequence of work history questions was not asked of younger workers.

---

[8] Results from the first stage of the project are discussed in more detail below, and in Linda Moeller (2002), cited above.

[9] Thombas B. Jabine and Fritz Scheuren (1985), "Goals for Statistical Uses of Administrative Records: The Next 10 Years," Journal of Business & Economic Statistics, 3 (4), pp. 380-391, and the comments that follow in the same volume: William P. Butz, "The Future of Administrative Records in the Census Bureau's Demographic Activities," pp. 393-395; John J. Carroll, "Uses of Administrative Records: A Social Security Point of View," pp. 396-397; Janet L. Norwood, "Administrative Statistics: A BLS Perspective," pp. 398-400; and Charles A. Waite, "The Future of Administrative Records in the Economic Programs of the Census Bureau," pp. 400-401.

[10] The SAS programs developed to implement these internal consistency checks are complicated, especially in the case of the SIPP. This complexity is evident in the flow charts that summarize the program that constructs the work experience variable. These flow charts describe the calculations undertaken to construct an actual work experience variable with data from the 1986-1992 panels. The sequence of work history questions included with the 1984 panel of the SIPP collected information on up to four jobs held at the beginning of the respondent's adult working life, and the code required to develop internal consistency checks was therefore more complex. The essential structure of the program is the same in both cases, however.

SIPP and SSA Work Experience Estimates, Employed Persons
Discrepancies by Cohort, Weighted Sample Observations

	Percentage by Cohort				Cohort
	SIPP < SSA	SIPP = SSA	SIPP > SSA	No Match	Share of Sample
Age	Females				
16-21	0.00	0.00	0.00	3.00	9.75
21-25	26.67	66.31	3.50	3.52	15.13
26-30	28.55	61.70	7.37	2.38	14.80
31-35	28.37	55.86	13.28	2.49	13.09
36-40	24.78	52.90	19.57	2.75	11.07
41-45	23.76	48.11	24.62	3.52	9.63
46-50	26.02	35.68	35.22	3.07	7.40
51-55	18.70	35.86	42.42	3.01	6.80
56-60	15.43	36.69	45.37	2.51	6.10
61-65	11.02	26.00	60.01	2.96	3.80
66-60	6.56	13.79	77.47	2.17	1.43
71+	6.49	8.07	83.46	1.98	0.99
Age	Males				
16-21	0.00	0.00	0.00	4.75	8.78
21-25	33.06	58.42	3.61	4.91	13.68
26-30	33.68	56.28	6.57	3.47	15.14
31-35	33.50	52.34	11.45	2.71	13.18
36-40	27.96	51.66	16.80	3.57	11.77
41-45	29.87	49.68	16.14	4.40	9.15
46-50	28.35	45.84	21.41	4.39	7.88
51-55	13.00	47.63	35.97	3.41	7.12
56-60	13.71	35.05	47.58	3.66	6.46
61-65	9.22	29.79	57.59	3.41	4.14
66-60	6.92	18.38	69.94	4.76	1.45
71+	0.00	7.07	92.30	0.63	1.25

Unweighted percentages of SIPP records that can be matched to SSA administrative records for the age group 16-21 are 87.52% among women, and 85.03% among men. Information on the duration of the current job, or last job held, is available, but the SIPP total work experience variable is not consistently available for persons in this age group. Therefore no comparison is made in this Table. Shares may not sum to 100%, due to rounding.

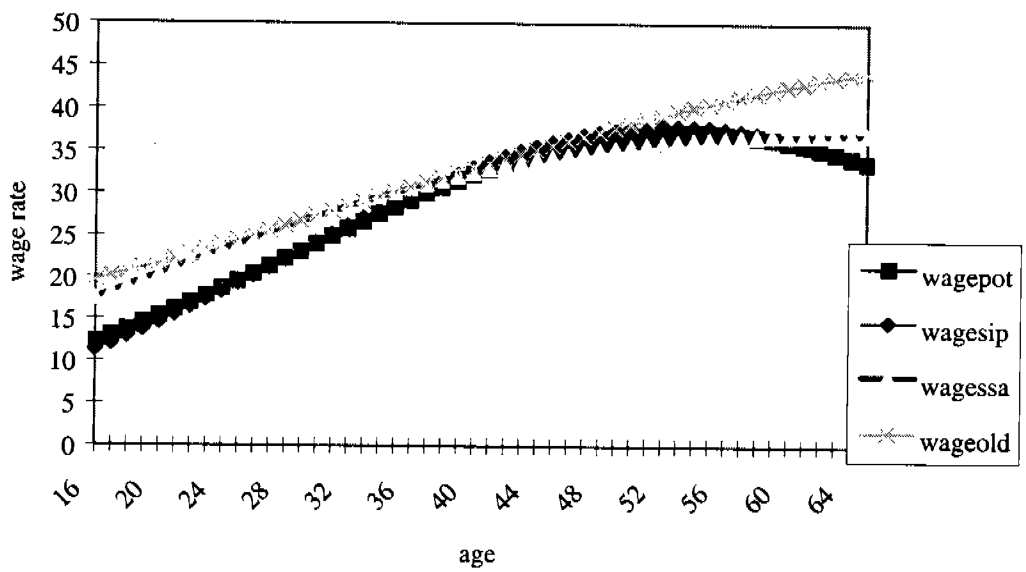
The bias in the SSA administrative record data seemed the more important for two reasons. First, prior OPT research with the NLSY found that accumulated work experience is not a significant determinant of young men's earnings. Similar results had previously been obtained by Mincer and Jovanovic in the case of young men, but they found actual work experience to be a significant determinant of the earnings of older men.[11] Second, while Honig and Hanoch have proposed econometric procedures to offset known variations in SSA coverage rates, a close accounting for the changes in SSA coverage rates over time is not possible because information on the industry and occupation in which respondents worked during the 1930's and 1940's is unavailable.[12] In sum, the use of SIPP work experience data seemed to be more the more straightforward approach.

[11] Jacob Mincer and Boyan Jovanovic (1981). "Labor Mobility and Wages." Studies in Labor Markets, S. Rosen, Ed., Chicago: University of Chicago Press.
[12] Also see Hanoch and Honig (1985), cited above.

Experience and earnings profiles were estimated separately with four alternative measures of actual work experience: a work experience proxy estimated with the coefficients of the 1973 CPS-SSA work experience equation that are provided in the labor composition bulletin, a work experience proxy estimated with SSA administrative record data matched to the 1984 SIPP, a work experience proxy estimated with information from the 1984 SIPP work history topical module, and potential work experience. In all cases examined, the 1973 and 1984 SSA experience profiles were flatter than the 1984 SIPP experience profiles, as would be expected when the average actual work experience of older workers is understated in the original data.

Perhaps more surprisingly, earnings profiles based on the 1973 CPS-SSA experience equation coefficients diverged from the earnings profiles obtained with the other three experience proxies, and this divergence increased with age. This result is illustrated in the chart below. In this chart the curve labeled *wagepot* corresponds to a wage equation in which potential experience is entered as a proxy for actual work experience. The profile obtained with a predicted value of the SIPP actual experience variable is labeled *wagesip*, the corresponding profile obtained with SSA data is labeled *wagessa*, and the profile obtained with the 1973 CPS-SSA proxy is labeled *wageold*. The pattern shown here, in which the earnings profile obtained with the 1973 CPS-SSA proxy is substantially higher than the other three curves in the case of older workers, was obtained for virtually all cases examined. In the case of the male wage profiles, wage profiles based on the 1973 proxy increase at an increasing rate.

Female Wage Rate Profile
SIPP and SSA Data
Ever Married, 17+ Years of Schooling, No Children



This finding suggests that combining 1973 experience equation coefficient estimates with 1984 earnings data results in biased wage rate estimates over age ranges in which the sample is thin.[13] It can be shown, by substituting the explanatory variables that enter the experience equation into the predicted and squared predicted experience values that enter the wage equation, that the work experience proxy implicitly

[13] This is like the outlier problem in time series data, but here the problem is due to groups of outliers, i.e., clusters of experience equation estimates that are too low, combined with the fact that the work experience equation and the wage equation have several explanatory variables in common.

introduces a number of structural restrictions between the work experience equation and the earnings equation. The results obtained with the 1973 CPS-SSA proxy and both the SSA and the SIPP proxies from 1984 appear to indicate that these implicit restrictions have not been stable over time. To the extent that the magnitudes of these discrepancies are not constant over time they may introduce a significant bias into the labor composition index because the weights associated with the growth rates of hours worked by older workers may be systematically overstated. The fact that the discrepancies seem to occur primarily over age ranges where the sample is thin suggests that the magnitude of the bias may be small. But if increased uncertainty about the real values of their retirement accounts induces members of the baby boom cohort to retire at older ages than their parents have done, as many analysts expect, then the relative importance of this bias may be sustained for a number of years to come.

Having determined that the 1984 SIPP data on actual work experience were stronger than the work experience measure available from SSA administrative record data, and that the incorporation of selection-bias correction factors into the estimation of the OPT wage equation was reasonably straightforward, the project moved into its second stage. An alternative strategy that was discussed then would have been to attempt to adjust for the changes in SSA coverage rates over time with statistical procedures similar to those discussed by Hanoch and Honig. But in light of the results just mentioned, which underscore the importance of working with microdata that are internally consistent from a temporal standpoint, it was decided instead to focus the second stage of the project on an examination of the stability of successive annual estimates of labor supply and wage equations with overlapping panels of the SIPP.

Similarly, at this juncture of the project some effort was made to generate adjusted t-statistics for the selection-bias corrected wage equation coefficients, following the calculations sketched in an appendix to the 1980 article by Heckman cited above, and to test the null hypothesis that the labor supply and wage equations represent a simply recursive system, following procedures outlined in the 1986 review article by Dhrymes.[14] However, development of the computer programs required to execute these calculations proved to be time consuming, and it is known that the second tests have not always been executed successfully in previous applications. Therefore further work to debug the SAS code that implements these tests was deferred until comparable results had been obtained with the SIPP panels that were fielded after 1984.

Summary

This appendix has reviewed results obtained during the initial stage of a major project to construct a new prototype labor composition index with a series of panels from the Survey of Income and Program Participation (SIPP). The SIPP is a new, nationally representative longitudinal household survey that is designed to provide accurate measures of actual hours worked and wage rates earned. Two independent measures of employees' total prior work experience, one based on responses to retrospective questions in the work experience topical module of the SIPP and one based on Social Security Administration (SSA) administrative records, were compared. This comparison revealed that the SSA estimates for older workers were biased downward due to low Social Security Insurance coverage rates at the inception of that program.

Current OPT procedures, in which a work experience proxy constructed with experience equation coefficients previously estimated with 1973 Current Population Survey (CPS) data matched to administrative records, were shown to generate biased wage rate estimates when incorporated into a 1984 wage equation. The OLS procedures applied in the construction of OPT's current labor composition index were then been modified to incorporate a standard selection-bias-correction factor. These calculations are much more complex than those required to work with the retrospective calendar-year microdata collected in the March CPS supplement. The construction of calendar-year values for earnings, hours worked, and wage rates requires re-indexing the SIPP's monthly variables, which are collected at four-month intervals,

---

[14] Phoebus J. Dhrymes (1986), "Limited Dependent Variables," Handbook of Econometrics Vol. III., Griliches and Intrilligator, Eds.

so that a given index number corresponds to the same calendar month for all rotation groups. Once the survey-month variables were re-indexed, annual estimates for the first calendar year spanned by the panel are obtained by summing or averaging monthly values 1-12, and 13-24. In light of this complexity, it is reassuring to learn that the wage equation estimates obtained with the SIPP are reasonably comparable to those obtained with annual microdata from the March CPS, for the same wage equation specification.

However, even after incorporation of a conventional selection-bias correction factor, earnings profiles based on the 1973 CPS-SSA matched file, and on the 1984 SIPP-SSA matched file, were found to be flatter than earnings profiles based on the actual work experience data reported in the SIPP work history topical module. That is, the SSA-matched files appear to over-predict the wage rates of younger workers and older workers. This finding suggests that the systematic under-estimates of employment among persons who have worked in industries not covered by Social Security legislation, and this effect is quite separate from the conventional selection bias problem discussed by Heckman. Surprisingly, a wage equation in which potential experience serves as a proxy for actual work experience resulted in earnings profiles that are quite comparable to those based on the SIPP actual work experience data, for all cases examined.

Appendix B: Identification

This appendix discusses the identification of the systems of equations presented in the body of the paper. The primary focus of the appendix is on the four-equation "structural" system of equations that includes the probability of living in a home owned by a household member. The other equations estimated are special cases of the four-equation system, except that the homeownership probit is omitted and the normalized asset income variable replaces the predicted probability of homeownership if the employment probit. Explanatory variables are listed in general terms below for simplicity; exact definitions of the variables employed in each system are provided in Appendix D.

Full identification of these systems requires taking explicit account of the fact that their stochastic elements are assumed to be distributed as a multivariate normal, and/or eliminating the assumption that the stochastic terms of the system have a normal distribution and relying on bootstrapping procedures to evaluate the structural restrictions imposed. However it seems worthwhile to note that, allowing for the normalization of coefficient values associated with the selection bias correction factor, it is possible to identify the coefficients of the four-equation system in which the wage rate, the probabilities of employment and living in a home owned by a household member, and accumulated work experience are determined simultaneously. Therefore, if identification and specification tests confirm that the model is identified and the stochastic terms of the system have been properly "scaled," and if the null hypothesis of no model misspecification is not rejected, we should be able to evaluate the relative importance of the explanatory variables that enter each equation in determining the values of the LHS variables.

The claim that the structural coefficients of the system can be identified is based on the following argument. Let $ex_{i,t}$ be the number of years worked experience accumulated by person i at time t. Let $h_{i,t}=1$ if respondent i lived in a home owned by a household member during calendar year t, and 0 otherwise. Let $e_{i,t}=1$ if reservation wage of respondent i falls below his or her market wage rate and 0 otherwise, as in the work of Heckman. Let $w_{i,t}$ be the log of the average hourly earnings of respondent i during calendar year t conditional on $e_{i,t}=I$, and let $ex_{i,t}$ be the total accumulated work experience at time t. Then the row vector $Y_{i,t}=\left(ex_{i,t}, h_{i,t}, e_{i,t}, w_{i,t}\right)$ consists of the current endogenous variables of the system.

It is assumed that the predetermined variables of the system, denoted $Y_{i,t-p}$, are years of work experience at $t-5$, education, the primary industry of employment, occupation, marital status, number of children, wealth, and geographic region of current residence. These variables are assumed to be predetermined in the sense that they change slowly due to substantial costs of adjustment. This assumption could be tested with endogeneity tests. The current exogenous variables of the system, denoted $X_{i,t}$, are assumed to be the wage of the spouse, a categorical variable that takes a value of 1 if the respondent reports a disability that makes it difficult to work and 0 otherwise, a categorical variable that takes a value of 1 if the last spell out of work was involuntary, and the age of the respondent.[1]

It is difficult to draw a sharp distinction between predetermined and exogenous variables in many instances. For example, in the case of long-lived marriages it might be argued that the wage of the spouse is a predetermined variable because the lifetime education levels and labor supply behavior of head and spouse have been optimized jointly. But in other cases this assumption seems unwarranted, given increases in divorce rates and dual earner households. Similarly, experience is a state variable with a large predetermined component, especially in the case of older workers.

In the current specification the predetermined variables included in $Y_{i,t-p}$ are assumed to be uncorrelated with $\varphi_{i,t}$. The stochastic component of the system is assumed to be distributed as a multivariate normal, $\varphi_{i,t}$ with a mean of zero and covariance matrix Φ. Under these assumptions the following logic shows that the system is formally identified.

The right-hand-side of the first equation given below is linear in the explanatory variables of the model, apart from $\varphi_{i,t}$.[2] The distribution of $\varphi_{i,t}$ is a known continuous function, with a known inverse, so G^{-1} exists. Applying this inverse to the first set of equations below, the identification of the linear, non-stochastic components of the system can be examined the in terms of the coefficient matrices on the right-hand side of the second set of equations.

$$Y_{i,t} = G\left[\alpha + Y_{i,t}B + \left(Y_{i,t-p}, X_{i,t}\right)C + \varphi_{i,t}\right]$$

$$\Rightarrow \quad G^{-1}\left(Y_{i,t}\right) = \alpha + YB_{i,t} + \left(Y_{i,t-p}, X_{i,t}\right)C + \varphi_{i,t}.$$

The structural restrictions of the system are the zeros in the matrices $B, C,$ and Φ. By analogy with general linear systems of econometric models (GLSEMs), this system is fully identified if the matrices composed of the explanatory variables that are omitted from each equation have full column rank.[3]

To see this more explicitly, write the vector of predetermined and exogenous explanatory variables of the system be written, $Z_{i,t}$, be written as follows:

[1] The number of years worked and age are "splined" to allow for changes in the probability of employment and home ownership with the approach of retirement and old age, respectively. These techniques are not pertinent to the identification conditions discussed here.

[2] For a more complete discussion of the identification problem as it arises in the estimation of hedonic equations, see Shulamit Kahn and Kevin Lang (1988), "Efficient Estimation of Structural Hedonic Systems," International Economic Review, 29, pp. 157-166, and the references cited in that article. In the current application, the use of categorical variables for industry and occupation, to control for industry-specific differences in human and physical capital employed in production, may have a rough correspondence to Kahn and Lang's discussion of variables that vary with the "matching process" in the case of multiple markets.

[3] Phoebus J. Dhrymes (1994), *Topics in Advanced Econometrics, vol. II: Linear and Nonlinear Simultaneous Equations*, New York: Springer-Verlag.

$$
\left(Y_{i,t-p}, X_{i,t}\right)' = \begin{pmatrix} ex_{i,t-5} \\ education \\ primary\ occupation \\ primary\ industry \\ marital\ status \\ number\ of\ children \\ wealth \\ geographic\ region \\ spouse's\ wage \\ disability \\ age \\ part-time \\ nowork \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \\ z_{10} \\ z_{11} \\ z_{12} \\ z_{13} \end{pmatrix} = Z_{i,t}',
$$

where $Y_{i,t} = \left(ex_{i,t}, h_{i,t}, e_{i,t} w_{i,t}\right)$. The structural restrictions of the system are the zeros in the following matrices:

$$
B = \begin{bmatrix} 0 & 0 & 0 & b_{1,4} \\ 0 & 0 & b_{2,3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \Phi = \begin{bmatrix} \sigma_{x,x} & 0 & 0 & \sigma_{x,w} \\ 0 & \sigma_{h,h} & \sigma_{h,e} & \sigma_{h,w} \\ 0 & \sigma_{e,h} & \sigma_{e,e} & \sigma_{e,w} \\ \sigma_{w,x} & \sigma_{w,h} & \sigma_{w,e} & \sigma_{w,w} \end{bmatrix}, \text{ and}
$$

$$
C = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} \\ 0 & c_{3,2} & c_{3,3} & c_{3,4} \\ 0 & 0 & 0 & c_{4,4} \\ c_{5,1} & c_{5,2} & c_{5,3} & 0 \\ c_{6,1} & c_{6,2} & c_{6,3} & 0 \\ c_{7,1} & c_{7,2} & 0 & 0 \\ 0 & c_{8,2} & c_{8,3} & c_{8,4} \\ 0 & 0 & c_{9,3} & 0 \\ c_{10,1} & c_{10,2} & c_{10,3} & 0 \\ c_{11,1} & c_{11,2} & c_{11,3} & c_{11,4} \\ 0 & 0 & 0 & c_{12,4} \\ c_{13,1} & 0 & 0 & 0 \end{bmatrix}
$$

The dependent variable in the first equation is years of work experience. The coefficients of the explanatory variables in that equation are represented by the first columns of the matrices B and C. All other endogenous variables in the system are omitted from $b_{.,1}$. The variables omitted from $c_{.,1}$ are occupation, industry, the wage of the spouse, and a categorical variable for part-time employment on the current job. The second equation pertains to the probability of living in a home owned by a household member; the coefficients of that equation are represented by the second columns of B and C. The other endogenous variables are omitted from the right hand side of that equation, as are occupation, industry, the wage rate of the spouse, and the categorical variable for full-time/part-time status.[4] However the probability of living in a home owned by a household member and the wage rate of the spouse are included in the third equation, while industry, occupation and full-time/part-time status are included in the fourth.

The matrix of coefficients that correspond to the explanatory variables omitted from the experience equation, $A_{1,o}$, is specified explicitly below. It is clear from inspection that the rank of $A_{1,o}$ is 3, unless the elements of that matrix are linear combinations of one another for some reason that is not evident.[5]

$$r\left(A_{1,0}\right) = r\left(\begin{array}{ccc} 0 & b_{2,3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ c_{3,2} & c_{3,3} & c_{3,4} \\ 0 & 0 & c_{4,4} \\ c_{8,2} & c_{8,3} & c_{8,4} \\ 0 & c_{9,3} & 0 \\ 0 & 0 & c_{12,4} \end{array}\right) = 3.$$

Similarly, in the cases of the homeownership and employment probits respectively:

---

[4] The current official methodology assumes that part-time status is an exogenous variable. However it is often argued that the decision to work part-time is an endogenous one, since women who work part-time allocate the balance of their work hours to child care and household production, and a substantial fraction of part-time workers are in school. For example, in 1998 roughly 27% of workers who usually work part-time reported "child care problems" or "other family or personal obligations" as their reason for working less than 35 hours per week, and roughly 34% cited "in school or training," or "retired or Social Security limit on earnings." The schooling and retirement decisions are clearly endogenous over the long run, and the same is usually true for number of children. But it is unclear *a priori* whether the part-time/full-time categorical variable should be considered predetermined or endogenous for the purposes of wage equation estimates. For the purposes of benchmarking the SIPP prototypes with the current official estimates thie exogeneity assumption has been maintained, but it should be tested in future work. The data cited here are annual estimates from Employment and Earnings, January 1999, Table 20.

[5] One of the functions of identification tests is to determine whether the matrices used to show that a system of equations can be identified do, in fact, meet these rank conditions when the coefficients of the system have been estimated with real-world data. The coefficient estimates obtained with econometric procedures are estimates of the "true" coefficients of the system, and the distribution of the coefficient estimates depends on the distribution of the stochastic terms of the true system. Identification tests take account of the fact that the matrices of omitted explanatory variables may be "almost" singular, without being singular in the non-stochastic sense of the word. Effectively, they are tests of the null hypothesis that the rank of the matrix $A_{j,o}$ is insignificantly different from 3.

$$r\left(A_{2,0}\right)=r\begin{pmatrix} 0 & 0 & b_{1,4} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c_{4,4} \\ 0 & c_{9,3} & 0 \\ 0 & 0 & c_{12,4} \\ c_{13,1} & 0 & 0 \end{pmatrix}=3, \text{ and } r\left(A_{3,0}\right)=r\begin{pmatrix} 0 & 0 & b_{1,4} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c_{4,4} \\ c_{7,1} & c_{7,2} & 0 \\ 0 & 0 & c_{12,4} \\ c_{13,1} & 0 & 0 \end{pmatrix}=3.$$

Finally, in the case of the wage equation we have: $r\left(A_{4,0}\right)=r\begin{pmatrix} 0 & 0 & b_{2,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ c_{5,1} & c_{5,2} & c_{5,3} \\ c_{6,1} & c_{6,2} & c_{6,3} \\ c_{7,1} & c_{7,2} & 0 \\ 0 & 0 & c_{9,3} \\ c_{10,1} & c_{10,2} & c_{10,3} \\ c_{13,1} & 0 & 0 \end{pmatrix}=3.$

Expanding the matrices $A_{j,0}$ to include the relevant submatrices of the covariance matrix Φ would not change the rank conditions shown above. Therefore the four-equation system is identified by conventional exclusion restrictions, assuming that these restrictions are appropriate.

Appendix C: Partitioning Considerations: Separability and Aggregation

Attribution of the sources of productivity growth to separate factors of production, i.e., to capital and labor, or to various categories of labor, requires that the aggregate production function be separable in the partitions within which factors of production are aggregated. For example, consider a "well-behaved" production function, $Q = F(K,L,E,M,S)$, where Q is real output, K represents the total capital stock, L represents total hours worked, E denotes energy, M represents total materials, and S represents total services. Each of these variables is a function of subsets of a large set of heterogeneous inputs $X = (X_1,\ldots,X_S)$ and outputs $Y = (Y_1,\ldots,Y_Q)$ that have been aggregated in some way, with weights that take account of relative value per unit, i.e, their relative price.[1]

Let lower case letters denote functions that are used to aggregate the individual inputs and outputs. In practice these functions tend to be arithmetic or geometric means of logs, levels, or square roots of the variables of interest.

$$\begin{aligned} K &= k(X_1,\ldots,X_K) \\ L &= l(X_{K+1},\ldots,X_L) \\ E &= e(X_{L+1},\ldots,X_E) \\ M &= m(X_{E+1},\ldots,X_M) \\ S &= s(X_{M+1},\ldots,X_S) \\ Q &= q(Y_1,\ldots,Y_Q) \end{aligned} \qquad (1)$$

Construction of multifactor productivity index numbers requires calculating the difference between the rate of growth of output and the share-weighted rates of growth of the inputs listed above. That is,

$$ln\, A_t - ln\, A_{t-1} \equiv (ln\, Q_t - ln\, Q_{t-1}) - (s_K (ln\, K_t - ln\, K_{t-1}) + \cdots + s_S (ln\, S_t - ln\, S_{t-1})).$$

This approach relies on the maintained hypothesis that the inputs $X = (X_1,\ldots,X_S)$ are separable in the partition (K,L,E,M,S). By definition, separability of labor from other factors of production is a situation in which:

$$\frac{\partial}{\partial X_k}\left(\frac{\frac{\partial F}{\partial X_i}}{\frac{\partial F}{\partial X_j}}\right) = 0, \quad \text{where} \begin{cases} i, j \in L \\ k \notin L \end{cases}, \quad I = Q,K,L,E,M,S. \qquad (2)$$

If the production function F is "strongly separable" in this partition, marginal change in the level at which factor X_k is employed has no effect on marginal productivity ratios among the various types of labor services included in the subset L. Separability of other factors is defined symmetrically.[2]

---

[1] This discussion abstracts from the issue of intermediate inputs. See Domar ().

[2] Complete definitions are provided in W. Leontief (1947), "A Note on the Interrelation of Subsets of Independent Variables of a Continuous Function with Continuous Derivatives," Bulletin of the American Mathematical Society, 55, pp. 343-350; and (1947), "An Introduction to a Theory of the Internal Structure of Functional Relationships," Econometrica, 15, pp. 361-373. Also see C. Blackorby and R. R. Russell (1989), "Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of Allen/Uzawa and Morishima Elasticities.)" American Economic Review, 79, pp. 882-888. Weak and strong separability are defined in Blackorby, Primont and Russell (1988, confirm), and I think in Berndt and Christensen.

To see how these conditions may be examined empirically, recall that the first-order conditions associated with simple models of profit-maximizing behavior on the part of perfectly or monopolistically competitive firms imply that workers will be employed at levels at which their relative wage rates are equal to the corresponding ratios of marginal products. That is, if input prices are denoted $w = (w_1, \dots, w_S)$, the traditional first-order conditions associated with the optimization of the firm's static (or time separable) objective function implies that factors are paid the value of their marginal products.

$$\frac{\frac{\partial F}{\partial X_i}}{\frac{\partial F}{\partial X_j}} = \frac{w_i}{w_j}, \quad i, j = 1, \dots, S. \tag{3}$$

Together the assumptions of (a) perfect or monopolistic competition and (b) separability of production functions imply that the relative wage rates among different categories of labor will be unaffected by an exogenous change in the level at which X_k is employed, because the ratio of marginal contributions that employees X_i and X_j make to total output will be unaffected:

$$\frac{\partial}{\partial X_k}\left(\frac{\frac{\partial F}{\partial X_i}}{\frac{\partial F}{\partial X_j}}\right) = \frac{\partial}{\partial X_k}\left(\frac{w_i}{w_j}\right) = 0, \quad \text{where } \begin{cases} i, j \in L \\ k \notin L \end{cases}. \tag{4}$$

In contrast with traditional general equilibrium theory, which describes a general situation in which "everything depends on everything else," these two assumptions allow for changes in the optimal *levels* at which factors within L, for example, are employed. But they imply there is no reason for the *composition* of the individual types of labor to change in response to a marginal change in X_k, because marginal productivity ratios and relative wage rates among these different categories of labor will be unaffected by an exogenous change in the equilibrium levels at which other factors in K,E,M and S are employed.

Since separability implies that the wage ratios in L are stable in the face of changes in relative prices among other factors of production, assign scalar values to each of the wage ratios in L, normalized to the L^{th} factor of production:

$$\begin{gathered}
\frac{\frac{\partial F}{\partial X_{K+1}}}{\frac{\partial F}{\partial X_L}} = \frac{w_{K+1}}{w_L} = a_{K+1,L} \Rightarrow w_{K+1} = a_{K+1,L} w_L \\
\vdots \\
\frac{\frac{\partial F}{\partial X_{L-1}}}{\frac{\partial F}{\partial X_L}} = \frac{w_{L-1}}{w_L} = a_{L-1,L} \Rightarrow w_{L-1} = a_{L-1,L} w_L .
\end{gathered} \tag{5}$$

In general the year-to-year difference between $X_{i,t}$ and $X_{i,t-1}$ will be small in magnitude, relative to the corresponding levels, $X_{i,t}$ or $X_{i,t-1}$. Therefore assume for simplicity that there is a set of scalars, $b_{K+1} \ldots b_L$, such that the following relationship holds over the two years t and $t+1$.

$$X_{i,t} = b_i X_{L,t}, \quad \text{where } i = K+1,\ldots,L. \tag{6}$$

These simplifying assumptions can be used to make the empirical content of the separability assumptions more explicit. To address the specific context in which the official BLS labor composition index is constructed, let the contribution of L to growth in total factor productivity between t and $t+1$ be measured with the following formula.

$$s_L(\ln L_t - \ln L_{t-1}) = s_L\{\ln[l(X_{K+1,t},\ldots X_{L,t})] - \ln[l(X_{K+1,t-1},\ldots X_{L,t-1})]\}. \tag{7}$$

$$\text{where } s_L = \frac{\left(\sum_{i=K+1}^{L} w_{i,t} X_{i,t}\right) + \left(\sum_{i=K+1}^{L} w_{i,t-1} X_{i,t-1}\right)}{\left(\sum_{i=1}^{S} w_{i,t} X_{i,t} + w_{i,t-1} X_{i,t-1}\right) + \left(\sum_{i=1}^{S} w_{i,t} X_{i,t} + w_{i,t-1} X_{i,t-1}\right)}.$$

Let the aggregator function l used to that maps the individual inputs $X_{K+1},\ldots X_L$ into the aggregate labor services L be arithmetic summation. Then we have:

$$l(X_{K+1,t}\ldots, X_{L,t}) = \sum_{i=K+1}^{L} X_{i,t} = \sum_{i=K+1}^{L} b_i X_{L,t} = X_{L,t}\sum_{i=K+1}^{L} b_i. \tag{8}$$

Substituting into (7) using the relationships in (5), (6) and (8), and assuming similar relationships also hold for sectors K, E, M and S, we obtain:

$$s_L(\ln L_t - \ln L_{t-1}) = \left[\frac{(w_{L,t} X_{L,t} + w_{L,t-1} X_{L,t-1})\sum_{i\in L} a_i b_i}{\sum_{j=K,L,E,M,S}\left[(w_{J,t} X_{J,t} + w_{J,t-1} X_{J,t-1})\sum_{i\in J} a_i b_i\right]}\right][\ln X_{L,t} - \ln X_{L,t-1}]. \tag{9}$$

The expression on the right-hand side of equation (9) implies that data from any arbitrary pair of variables (w_L, X_L) in L captures the true behavior of all elements (w_i, X_i) in L at between *t-1* and *t*, when weighted by a scalar equal to $\sum_{i=K+1}^{L} a_i b_i$. Thus the expressions in equations (5) and (6) convert a very general analytical framework with many heterogeneous factors of production into one that is comparable to traditional models in which factors of production within each major category of inputs "move together."

When these assumptions are justified, the measurement of the effect of a change in X_k on the contribution that factors in L make to total productivity growth is relatively straightforward. An exogenous change in $X_k, k \notin L,$ has no effect on the numerator of the expression within square brackets on the right hand of equation (10), That is, equations (5) and (7) imply:

$$\frac{\partial}{\partial X_k}\left[\frac{\frac{\partial F}{\partial X_i}}{\frac{\partial F}{\partial X_L}}\right]=\frac{\partial}{\partial X_k}\left[\frac{w_i}{w_L}\right]=\frac{\partial}{\partial X_k}a_{iL}=0,$$
$$\frac{\partial}{\partial X_k}\left[\frac{X_i}{X_L}\right]=\frac{\partial}{\partial X_k}b_{iL}=0. \tag{10}$$

Therefore the effects of an exogenous change in X_k on total factor productivity growth are channeled entirely through the expressions like $(w_{J,t}X_{J,t}+w_{J,t-1}X_{J,t-1})$ and $(ln\,X_{J,t}-ln\,X_{J,t-1})$. The equilibrium employment levels and factor price levels of all the heterogeneous inputs within a given sector respond proportionately to a given change, according to the relationships given by equations (5) and (7). These assumptions clearly support the use of probability samples from each major sector to measure changes taking place throughout the entire economy.

In the case of labor services, it has often been observed that employment in capital intensive industries tends to be more cyclical than employment in industries that experience less cyclical demand fluctuation. Similarly, the determinants of demand fluctuations in the trade industries may be different from those in the service industries. In such cases the values of the scalars $b_{K+1}\ldots b_L$, defined above equation (7), may be significantly more stable *within* subsets H,T,S than they are across the entire set of labor services.

Intuitively, the simplifying assumption that equilibrium employment levels for different categories of labor services can be simply scaled up or down to obtain a reliable measure of the total may be more palatable for subsets of labor services, such as the heavy industry, trade, and service industries, than it is for all labor services supplied to the market.[3] Then the contribution of labor services to output and multi factor productivity growth might be captured more accurately as a weighted sum of the labor services supplied by these three categories of labor. The labor composition index takes account of these systematic variations in the characteristics of labor services.

Following this line of argument, let the individual inputs that are categorized as labor services employed in production , denoted L above, be subdivided further into three major industry subsets: hours worked in heavy industry, trade and finance, and services.

$$\begin{aligned} L_H &= h(X_{K+1},\ldots,X_H) \\ L_T &= t(X_{H+1},\ldots,X_T) \\ L_S &= h(X_{T+1},\ldots,X_S) \end{aligned} \tag{11}$$

Let the contribution of labor services to multifactor productivity growth be measured with the following expression:

$$s_L(ln\,L_t-ln\,L_{t-1})=s_H(ln\,L_{H,t}-ln\,L_{H,t-1})+s_T(ln\,L_{T,t}-ln\,L_{T,t-1})+s_S(ln\,L_{S,t}-ln\,L_{S,t-1}), \tag{12}$$

where: $$s_{i,t}=\frac{\left(\sum_j^{j'} w_{i,t}X_{i,t}\right)+\left(\sum_j^{j'} w_{i,t-1}X_{i,t-1}\right)}{\sum_{i=1}^{L} w_{i,t}X_{i,t}+w_{i,t-1}X_{i,t-1}},$$

$$ln\,L_{i,t}-ln\,L_{i,t-1}=(ln\,X_{i,t}-ln\,X_{i,t-1})\sum_j^{j'}a_i b_i \text{ , and}$$

$$i,j'=H,T,S,\quad j=K+1,H+1,T+1,$$

[3] Theil, Linear Aggregation.

In this case, large cyclical fluctuations in wage and employment levels, which are characteristic of heavy industry, are contained within the sector that is experiencing them, rather than being averaged out. This is particularly noteworthy in the case of overtime wage rates, which command a premium but seem more attributable to adjustment costs than to the skill characteristics of workers.

In real applications, of course, a single observation is not taken to be representative of a single sector, or a single cell. Instead, complex stratified survey data are used to calculate weighted mean or conditional mean wage rates and totals for hours worked, $(\overline{w}_J, X_J)$, where the index J ranges over the cells in which the survey microdata are partitioned. Wage rates in heavy industry tend to have higher mean values and to be more highly skewed than wage rates overall, cyclical variations in average wage rates are likely to be more pronounced when the data are not partitioned by industrial sector. Therefore the assumption that labor services within the partitions H,T, and S retain a proportionate relationship to one another over the course of the business cycle may be more acceptable than the broader partition in the case of labor services.

The separability assumptions discussed above are imposed implicitly when it is assumed that the relationship between outputs and inputs can be represented by a Cobb-Douglas or constant elasticity of substitution (CES) production function.[4] Systems of factor demand equations derived from more flexible functional forms, such as the translog and the generalized Leontieif, have been used to test these assumption.[5]

---

[4] Hirofumi Uzawa (1962), "Production Functions with Constant Elasticity of Substitution," Review of Economic Studies, 29, pp. 291-299; Ernst R. Berndt and Laurits R. Christensen (1973), "Internal Structure of Functional Relationships: Separability, Substitution, and Aggregation," Review of Economic Studies, 40, pp. 403-440.

[5] To my knowledge, these tests were first presented in the literature by Christensen, Jorgenson and Lau in their 1971 paper, "Conjugate Duality and the Transcendental Logarithmic Production Function," Econometrica, 39. Some limitations of this approach are illustrated in Guilkey, Lowell, and Sickles (1983), "Comparison of the Performance of Three Flexible Functional Forms," International Economic Review, 24, pp. 591-616, and R. E. Lopez (1985) "Structural Implications of a Class of Flexible Functional Forms for Profit Functions," International Economic Review, 26, pp. 593-601.

Appendix D: Variable Definitions and Estimating Equations

$$pot = (age - schooling - 6)$$

$$priv = (1 \quad \text{if currently employed in nonprofit industry}, \quad 0 \text{ otherwise})$$

$$school = \begin{pmatrix} s0to4 \\ s5to8 \\ s12 \\ s13to15 \\ s16 \\ s17up \end{pmatrix} = \begin{pmatrix} 1 \quad if\ schooling \in [0,4], \quad 0\ otherwise \\ 1 \quad if\ schooling \in [5,8], \quad 0\ otherwise \\ 1 \quad if\ schooling = 12, \quad 0\ otherwise \\ 1 \quad if\ schooling \in [13,15], \quad 0\ otherwise \\ 1 \quad if\ schooling = 16, \quad 0\ otherwise \\ 1 \quad if\ schooling \geq 17, \quad 0\ otherwise \end{pmatrix}$$

$$kids = \begin{pmatrix} kid1 \\ kid23 \\ kid4+ \end{pmatrix} = \begin{pmatrix} 1 \quad if\ kids = 1, \quad 0\ otherwise \\ 1 \quad if\ kids = 2,3, \quad 0\ otherwise \\ 1 \quad if\ kids \geq 4, \quad 0\ otherwise \end{pmatrix}$$

$$race = \begin{pmatrix} black \\ hisp \end{pmatrix} = \begin{pmatrix} 1 \quad if\ black, \quad 0\ otherwise \\ 1 \quad if\ hispanic, \quad 0\ otherwise \end{pmatrix}$$

$$married = (1\ if\ ever\ married, 0\ otherwise)$$

Additional explanatory variables for alternative specifications are as follows:

$$schooling = years\ of\ school\ completed$$

$$aged = \max(pot - 55, 0)$$

$$spled = \begin{pmatrix} spled \\ spledhs \\ spledsc \\ spledcd \\ spledgd \end{pmatrix} = \begin{pmatrix} schooling \\ \max(schooling - 11.5, 0) \\ \max(schooling - 12.5, 0) \\ \max(schooling - 15.5, 0) \\ \max(schooling - 17.5, 0) \end{pmatrix}$$

$$assypov = \left({}^{family\ income\ from\ financial\ assets} \big/ {}_{family\ poverty\ cutoff\ value} \right)$$

$$spouse = \begin{pmatrix} relage \\ reled \\ othwage \end{pmatrix} = \begin{pmatrix} own\ age / age\ of\ spouse \\ own\ schooling / schooling\ of\ spouse \\ wage\ rate\ of\ spouse \end{pmatrix}$$

$$duroutex = \begin{pmatrix} 1\ if\ last\ spell\ of\ no\ work\ \geq 6\ months\ was\ involuntary, \quad 0\ otherwise \\ duration\ last\ spell\ of\ no\ work \geq 6\ months\ if\ involuntary, \quad 0\ otherwise \end{pmatrix}$$

$$disab = (1\ if\ disabled, \quad 0\ otherwise)$$

$$ind = \begin{pmatrix} hvy \\ tnf \\ svs \end{pmatrix} = \begin{pmatrix} 1\ if\ \max\left({}^{industry\ hours\ worked} \big/ {}_{total\ hours} \right) \in heavy\ industry, \quad 0\ otherwise \\ 1\ if\ \max\left({}^{industry\ hours\ worked} \big/ {}_{total\ hours} \right) \in trade\ and\ finance, \quad 0\ otherwise \\ 1\ if\ \max\left({}^{industry\ hours\ worked} \big/ {}_{total\ hours} \right) \in services, \quad 0\ otherwise \end{pmatrix}$$

$$ms = (1\ if\ currently\ married, \quad 0\ otherwise)$$

$$young = number\ of\ children\ with\ age \leq 6$$

$$spage = age\ of\ spouse\ if\ married, \quad 0\ otherwise$$

$$sped = schooling\ of\ spouse\ if\ married, \quad 0\ otherwise$$

$$othwage = wage\ of\ spouse\ if\ married\ and\ spouse\ is\ employed, \quad 0\ otherwise$$

$$long = \begin{pmatrix} \mathrm{int} \ln g \\ wkd \ln g \end{pmatrix} = \begin{pmatrix} 1 \quad if\ tenure \geq 10, \quad 0\ otherwise \\ (\max(pot\exp - tenure, 0)) \quad if\ tenure \geq 10, \quad 0\ otherwise \end{pmatrix}$$

$$short = \begin{pmatrix} \mathrm{int}\ yng \\ wkdyng \end{pmatrix} = \begin{pmatrix} 1 \quad if\ age \in [16,21), \quad 0\ otherwise \\ (\max(pot - tenure, 0)) \quad if\ age \in [16,21), \quad 0\ otherwise \end{pmatrix}$$

$$invpot = (pot\exp)^{-1}$$

$$last5 = worked\ at\ t-5$$

Dependent variables for experience equation

$$qtrswkd = total\ quarters\ of\ employment\ recorded\ in\ SSA\ records$$

$$worked = number\ of\ years\ employed\ 6+ months\ as\ of\ year\ t$$

$$\ln wkd = (\log(worked)\quad if\ worked > 0,\quad NA\ otherwise)$$

Alternative Experience Equations Specifications

The experience equation currently employed by the BLS is specified separately for males and females, as identified by subscripts *f* and *m* below:

$$qtrswkd_f = \lfloor school, pot, married, kids, pot^2, (pot, pot^2) * married, pot * (kids, school, priv), priv \rfloor$$

$$qtrswkd_m = \lfloor school, pot, pot^2, pot * (school, priv), priv \rfloor$$

Alternative experience equation following current BLS approach, but with the same specification estimated separately by gender:

$$worked = \lfloor school, pot, married, kids, pot^2, (pot, pot^2) * married, pot * (kids, school, priv), long, short \rfloor$$

Alternative experience equation following Heckman (1980), and where $x \bullet x$ identifies "interactions of all linear terms" in the vector x . The variable *assypov* replaces assets in Heckman's work. The same specification is estimated separately by gender:

$worked = \lfloor young^2, x \bullet x \rfloor$, where $x = (young, assypov, reled, relage, othwage, schooling)$.

Alternative long-run structural specification, following the work of Mincer and Becker in which "gains from trade" from marriage are possible. A spline function in years of schooling replaces the quadratic in potential experience and the schooling dummies; the intervals used to define the spline are the same as those used to define the schooling dummies. The duration of the last spell out of work of 6 months or longer, when that spell was identified as a period when the respondent was unable to find work, is assumed to be exogenous variable that will be negatively correlated with the total number of years worked. A reported disability is also assumed to be exogenous and negatively correlated with years worked. The same specification is estimated separately by gender:

$$worked = \lfloor pot\exp, pot\exp^2, spled, kids, kids^2, spouse, assypov, duroutex, disab, aged, race \rfloor$$

Alternative specification in which the dependent variable is transformed to the logarithm of years worked, following Lancaster and Chesher:

$$\ln wkd = \lfloor \ln pot, \ln pot^2, spled, kids, kids^2, spouse, assypov, duroutex, disab, aged, race \rfloor$$

Alternative short-run specification in which accumulated work experience in year is assumed to be a predetermined state variable. Results are reported separately by gender , as above:

$$worked = \lfloor last5, last5^2, spled, young, young^2, spouse, assypov, duroutex, disab, aged, race \rfloor.$$

Alternative functional form specification with an "S-curve," employed in models of market saturation:

$$\ln wkd = \lfloor invpot, spled, young, young^2, spouse, assypov, duroutex, disab, aged, race \rfloor$$