ECONOMICS BULLETIN

A quality adjusted wage index

Jennjou Chen
National Chengchi University

J. S. Butler
University of Kentucky

Abstract

In this paper, a new method of estimating a wage index is proposed and implemented. We construct a wage index by controlling for quantity, as well as quality of labor. Our approach uses a set of year dummies as the basis for calculation of a wage index. The March Current Population Survey Supplement (1983-2000) is employed, and empirical wage equation models are estimated in this paper. The estimation results of the proposed wage index suggest that the existing Employment Cost Index perhaps overestimates the increases in wages adjusted for quality.

Corresponding author: Jennjou Chen, Department of Economics, National Chengchi University, Wenshan, Taipei City 11605, Taiwan, phone number: 886-915-045-248, email: jc171@cornell.edu. Thanks are due to Tsui-Fang Lin for her valuable comments on an earlier draft of this paper. The authors alone are responsible for errors and opinions.
Citation: Chen, Jennjou and J. S. Butler, (2006) "A quality adjusted wage index." *Economics Bulletin*, Vol. 10, No. 13 pp. 1-14
Submitted: October 8, 2006. **Accepted:** November 13, 2006.
URL: http://economicsbulletin.vanderbilt.edu/2006/volume10/EB-06J30002A.pdf

I. Introduction

A price index is a composite indicator of price changes, used to compare the price of a commodity or service across periods. A price index may represent changes in the price of a particular commodity or of several commodities, controlling for quantitative composition (in case of several commodities) and quality in the base period. Several price indices represent the overall changes in economy-wide price levels. The Consumer Price Index (CPI) is among the most popular and widely used indices, though economists often argue that the CPI overestimates true increases in cost of living and does not reflect the changes in the quality of consumption of goods (Bils and Klenow, 2001).

There is one wage-related price index reported by the United States Bureau of Labor Statistics (BLS): the Employment Cost Index (ECI). The ECI measures quarterly and annual changes in labor costs. Only the key features of the ECI are summarised in this section; details can be found in Ruser (2001) and at the BLS web home page.

The ECI might be a good index of changes in employers' labor costs but perhaps it is not a very 'reflective' wage index. In order to obtain a desirable wage index, the quantity and the quality of labor in the base period have to be controlled. The quantity of labor is probably controlled in the ECI but not the quality aspect. For instance, higher average years of schooling or higher average years of work experience could contribute to increase in the average wage rate. If we did not control for the schooling variable, the effect of schooling on wages would be counted as part of the wage change, and this would cause an overestimation problem when constructing a wage index.

The main goal of this study is to develop a good wage index, including the relevant methodology, that represents average changes in wage rates. We suggest a new approach to estimating the yearly wage index, one that represents the overall economy-wide changes in wage rates. The idea behind our approach is mainly based on the quality aspect. We will try to control for all the important factors that affect wage determination, such as sex, education, work experience and age, in our wage index model. In order to control for these variables, a good household-based survey data set might be needed because it is uncommon to have quality related variables in an employer-based survey like the NCS. In this paper, we will use data from the March Current Population Survey Supplements (1983-2000), which is among the best and the most commonly used household-based survey data sources in the United States.

A typical wage equation estimation technique will be performed and year dummies will be taken into account in our model (Killingsworth, 1983). The yearly wage index will be derived, based on the estimated coefficients of the year dummies. In section 2, the theoretical basis for this paper will be discussed. In section 3, the data used in estimation will be examined. The estimation results are reported in Section 4, and the conclusion is summarised in Section 5.

II. The Model

The study of the wage equation has probably been the single most popular subject in the field of labor economics during the past 30 years. However, less attention has been paid by researchers to investigation of the wage index. In this section, we will use the estimated coefficients for the year dummies in the wage equation to derive a general wage index.

In our model, we assume that the structure of the wage equation in every year is the same; so the wage differences in different years come primarily from the overall wage changes. All the unexplained differences of wages across years are the year effects. Therefore, we can use the year effects to calculate a wage index. However, this assumption might be too strong. For example, a stable wage equation requires that the relationship between schooling and earnings is constant over time. Existing literature (Katz and Murphy, 1992, and Murphy and Welch, 1989) argues that the college wage premium might change over time. Thus, the proposed wage index could not only have the real wage change effects, but it could also include non-constant college wage premium effects, which would cause the wage index to be biased. In the following discussion, a stable wage equation structure will be assumed, in order to simplify the calculation of the proposed wage index. Then, this stability assumption will be relaxed, and the potential problems and solutions will be addressed in greater detail at the end of this section.

Now consider that we have the following typical log wage equation:

$$ln\,[W_t\,/\,(1+i_t)] = X_t'\beta + \varepsilon_t, \qquad (1)$$

where $ln(.)$ is the natural log function, i_t is the degree of wage inflation since the baseline year, X_t is a vector of variables which can explain wages at time t, and ε_t, is the disturbance at time t. Under this wage equation setup, we implicitly assume that once we discount the wages at period t by $1 + i_t$, the structure of the wage equation stays the same, i.e. the equation is linear and β is stable. Let us move $(1 + i_t)$ to the right hand side, and Equation (1) becomes

$$ln\,(W_t) = C_t + X_t'\beta + \varepsilon_t, \quad t = 1, 2, 3,\ldots, T, \qquad (2)$$

where $C_t = ln\,(1 + i_t)$. Thus, C_t is the measure of pure changes in wages for different periods.

In addition to the pure changes in nominal wages (i.e. the C_i), the term $X_t'\beta$ can be decomposed into two parts: the price effects and the quality effects. First, some of the potential covariates might affect the wages directly, but not the quality of the workforce. Union status, gender dummy, and metropolitan area dummy are examples of covariates with price effects. Acquisition of human abilities through formal education and on-the-job experience, on the other hand, affect the quality of labor. Thus, the effects that years of education and work experience have on wages are quality effects that should properly be removed from an index of wages. Equation (2) can then be re-written as follows:

$$ln\,(W_t) = (C_t + X_{p,t}'\beta_p) + X_{q,t}'\beta_q + \varepsilon_t, \quad t = 1, 2, 3,\ldots, T, \qquad (3)$$

where $X_{p,t}'\beta_p$ are the price effects, $X_{q,t}'\beta_q$ are the quality effects, and $\beta = [\beta_q, \beta_q]$. As a result, $(C_t + X_{p,t}'\beta_p)$ measures the wage changes in different periods, controlling for change in quality of the labor force (i.e. $(C_t + X_{p,t}'\beta_p) = ln\,(W_t) - X_{q,t}'\beta_q - \varepsilon_t$, and it is the wage adjustment which eliminates changes caused by quality from the actual overall wages).

We can generate the year dummies for every year and estimate equation (3) in the Ordinary Least Square (OLS) model. After we obtain the estimates of $(C_t + X_{p,t}'\beta_p)$ for $t = 1, 2, 3,\ldots, T$, the wage index is equal to

$$W_t/W_b = exp(C_t + X_{p,t}'\beta_p)\,/\,exp(C_b + X_{p,b}'\beta_p) \qquad (4)$$

for year t, and equal to one for the base year b. The delta-method can be applied to obtain the variance of the wage index.

Because of the stability assumption on the wage equation, the proposed wage index might produce biased estimates for the true changes in labor costs. The coefficients are assumed to be the same for all periods, i.e. $[\beta_q, \beta_q]$ is a vector of constants and it does not change over time. If this stability assumption is not true, $[\beta_q, \beta_q]$ needs not be the same for all periods. Since $[\beta_q, \beta_q]$ is forced to remain unchanged over time when we estimate the above wage equation, the estimated $[\beta_q, \beta_q]$ will be biased. Thus, the proposed wage index will not estimate the true changes in labor costs.

This problem could be solved by using a new set of interaction terms between the covariates and the year dummies. Let us add the interaction terms into equation (2); it becomes

$$ln\,(W_t) = C_t + X_t'\beta + X_t'\alpha_t + \varepsilon_t, \quad t = 1, 2, 3,\ldots, T, \qquad (5)$$

where $X_t'\alpha_t$ is the interaction term. This equation can be re-written as

$$ln\,(W_t) = C_t + X_t'\beta_t + \varepsilon_t, \quad t = 1, 2, 3,\ldots, T, \qquad (6)$$

where $\beta_t = \beta + \alpha_t$. Thus, by adding these interaction terms, the coefficients are allowed to vary over time.

III. The Data

A good household-based survey data set is needed in order to obtain some quality related variables. The Current Population Survey (CPS) is a good data source. We use data from the Current Population Survey Annual Earnings Files (CPSAEF), produced by the National Bureau of Economic Research (NBER), in the United States. The CPSAEF has data from 1979 to 2000 but some key variables are missing in the early years. For example, information about labor union status before 1983 is not available. So, we only use data from 1983 to 2000. The sample size for this analysis is 5,463,536. One potential advantage of using such a large data set is to obtain much more reliable estimates for the wage index.

The key variable in this study is the wage, EARNWKE. There are two types of wage information reported in the CPSAFE: hourly wage and weekly wage. We choose to use the weekly wage rate because this will avoid the measurement error problem in the reported weekly hours.[2] Among the 5,463,536 respondents, there are 2,940,849 with positive wage rates and 2,522,687 without wage information.

YEAR84–YEAR00 are year dummy variables. Year 1983 is the base year and is not in the set of year dummies. SEX is a dummy variable which equals one if the respondent is male. AGE is the respondent's age. UNION is a dummy variable that equals one if the respondent is a union member or is otherwise covered by a union or employee association contract.

EDUCATION is respondent's years of schooling. EXP and EXP2 are potential work experience and its square, respectively. The potential work experience is calculated by AGE – EDUCATION – 6. This same formula is employed in most empirical wage equations that lack an experience variable in the data set. WHITE is a dummy variable equal to one if the respondent's race is white. SMSASTAT is a dummy variable equal to one if the respondent lives in a metropolitan area. MARITAL is a dummy variable equal to one if the respondent is married. FT is a dummy variable equal to one if the respondent works full-time (35 or more hours a week).

All above covariates are interacted with the sex dummy, so that the estimated coefficients could be different by gender. The variables' names beginning with 'M_' are the male interaction terms, and the ones with 'F_' are the female interaction terms. Also industry and occupation dummies are used in all models in order to control the industry and occupation differences. The summary statistics of all the variables are reported in Table 1.

IV. The Results

We assume the stability of the wage equation for simplicity, so that, instead of model (6), the model for equation (3) is estimated. The estimated wage index results from the following two methods will be presented:

1. OLS with only year dummies,
2. OLS with all potential explanatory variables,

The estimation results are shown in Table 2. All coefficients are within a reasonable range, and they are all statistically significant at the 1 percent level. These estimates will be employed to compute the proposed quality adjusted wage indices, based on the formula in equation (4). Notice that the quality effects from an individual's years of schooling and potential work experience are removed from the estimated wage indices, except the one using OLS, with only year dummies.

The estimated wage indices from OLS with only year dummies and from OLS with all potential explanatory variables are shown in Table 3. CPI[1], ECI[2], and GDP Deflator (GDPD)[3] are also included in Table 3 for the purpose of comparison. We find that the wage index obtained from the OLS model with only year dummies is closely related to all three price indices (i.e. CPI, ECI and GDPD), particularly to the ECI. We find that the wage index and ECI are basically next to each other in Figure 1. This makes perfect sense since the ECI is computed using an idea similar to the wage index, but without controlling for other potential explanatory variables related to the quality of labor. The correlation coefficients between the wage index and each of the three price indices are 0.994 (CPI), 0.999 (ECI), and 0.987 (GDPD).

The wage index from OLS with year dummies and all other potential explanatory variables is graphed in Figure 2. Notice that the wage index becomes much lower than the one in the OLS model with only year dummies. Once we add some explanatory variables to control for the quality of labor, in addition to the year dummies in the OLS model, we have a smaller year effect. This implies that if we fail to control for other potential explanatory variables to explain the quality of labor when estimating the wage index, we are more likely to overestimate the true wage increase.

The differences between the empirical estimates of the ECI and our proposed indices are large over the period 1983–2000, the proposed index increasing by about 73.5 per cent against an increase of 84.3 per cent in the ECI. In addition, the ECI is larger in all years except 1984. Interpretation of macroeconomic and labor market conditions could be affected by such differences.

Finally, Table 4 presents the summary statistics of some variables (1983–2000). These variables are EARNWKE, AGE, EXP, EXP2, and EDUCATION. The first column is the yearly mean of weekly earnings for years 1983 to 2000. The trend of EARNWKE increases over time. The wage index using only the information of EARNWKE is exactly the one that we obtain from the OLS model with only the year dummies. Part of the year effects will be due to the upward trend in the quality of the workforce in terms of the human capital variables (work experience (EXP and EXP2) and years of schooling (EDUCATION)). This is reflected in the declining impact of the year dummies in the final three columns of Table 2. For instance, the potential work experience (EXP) increases from year 1983 to 2000. In general, the longer the work experience workers have, the higher the wage rates they will obtain because work experience will reflect human capital acquired on the job. Consequently, the increase in the average wage rate from 1983 to 2000 is in part due to a higher average work experience in the labor force. Longer years of schooling could be another reason for the increase in the average wage rate. As a result, we have to control for the quality of the labor force, in order to obtain a proper wage index.

V. Conclusion

In this paper, we propose an alternative approach to estimating a wage index. The major contribution of this paper is that we estimate various sets of wage indices, and take into account the changes in the quality of the labor force. The year dummies are used to represent the major economy-wide changes in wages, after controlling for other social and economic factors. The

[1] The CPI data can be downloaded from the Bureau of Labor Statistics' web site under the following link: ftp://ftp.bls. gov/pub/special.requests/cpi/cpiai.txt.

[2] The ECI data can be downloaded from the Bureau of Labor Statistics' web site: ftp://146.142 .4.23/ pub/suppl/ECI.ECHISTRY.TXT. The wage and salary ECI is used for the comparison because we use wage and salary to create our wage index.

[3] The GDPD data can be downloaded from the Bureau of Economic Analysis' web site: http:// www.bea.doc.gov/bea/dn/nipaweb/TableViewFixed.asp#Mid.

results of the estimated wage indices suggest that the ECI might overestimate the true changes in wages.

A wage index controlling for the quality of labor gives us a more accurate estimation of the changes in wages and it may have a number of practical uses. Monetary and fiscal policy-makers need a more accurate measure of the actual changes in employers' labor costs; changes caused by quality improvement should not be counted as part of labor cost changes. Also, a quality adjusted wage index can serve as a better inflator in collective bargaining contracts and long-term purchasing contracts. Furthermore, a quality adjusted wage index can be used to discount the nominal wage, and will produce more accurate results when we estimate panel data models involving wage rates.

In the future, we could control for other factors in the wage equation model to make the wage index closer to the actual changes in wages. For example, a panel data model controlling for individual effects should be a better model to estimate the wage index since the individual effects might explain a large part of wage variation within the individual's wage rates across periods. Also, we might want to use linked employer-employee data sets in order to control the firm effects, because the firm effects play an important role in wage determination (see Abowd and Kramarz, 1999).

Reference

[1] Abowd, J. M. and F., Kramarz, (1999), "*The Analysis of Labor Markets Using Matched Employer-Employee Data*," in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, Volume 3B, Chapter 40, 2629-2710, Amsterdam: North Holland.

[2] Bils, M, and J. Klenow, (2001), "*Quantifying Quality Growth,*" American Economic Review, 91, 1006-1030.

[3] CPS Merged Outgoing Rotation Groups CD-ROM (2000), National Bureau of Economic Research.

[4] Employment Cost Indices 1975-99 (2000), Bureau of Labor Statistics, September.

[5] Heckman, J., (1979), "*Sample Selection Bias as a Specification Error*," Econometrica, 47, 53-162.

[6] Katz, L.F. and K.M. Murphy, (1992), "*Changes in relative wages, 1963–1987: supply and demand factors*," Quarterly Journal of Economics, 107, 35–78.

[7] Killingsworth, M. R., (1983), *Labor Supply*, Cambridge: Press Syndicate of the University of Cambridge.

[8] Murphy, K. and F. Welch, "*Wage premiums for college graduates. Recent growth and possible explanations,*" Educational Researcher, 18, 17–26.

[9] Ruser, W. J., (2001), "*The Employment Cost Index: what is it?*" Monthly Labor Review, 124, 3-21.

Table 1: Summary Statistics (N=5,463,536)

Variables	Definition	Working (N=2,940,849) Mean	Working (N=2,940,849) Standard Deviation	Not Working (N=2,522,687) Mean	Not Working (N=2,522,687) Standard Deviation
EARNWKE	For a hourly paid worker, this is his/her hourly earnings time his/her weekly hours. For an weekly paid worker, this is his/her weekly wage.	445.25	341.58	.	.
YEAR84-YEAR00	Indicator variables which equal one if respondent was in that year.	.	.	.	.
SEX	Indicator variable which equals one if respondent was male.	0.51628	0.49974	0.41592	0.49288
AGE	Respondent's age.	37.651	12.892	50.390	21.092
UNION	Indicator variable which equals one if respondent was a union member or covered by a union or employee association contract.	0.16046	0.36703	0.00017	0.01319
EXP	Potential work experience which is computed by age - schoolling - 6.	18.379	13.119	32.495	21.742
EXP2	Potential work experience square.	509.90	616.40	1528.66	1485.02
SCHOOLLING	Years of schoolling	13.295	2.6028	11.926	3.154
WHITE	Indicator variable which equals one if respondent's race was white.	0.86171	0.34520	0.85352	0.35358
SMSASTAT	Indicator variable which equals one if respondent lived in a metropolitan area.	0.75796	0.42832	0.70643	0.45540
MARITAL	Indicator variable which equals one if respondent was married.	0.59806	0.49029	0.57725	0.49400
FT	Indicator variable which equals one if respondent worked full-time (35 or more hours/week).	0.79425	0.40425	.	.

Sources: Annual Earnings Files of the 1983 to 2000 Current Population Survey, Bureau of Census.

Table 2: Estimation Results for Wage Equations

Variables	OLS With Only Year Dummies	OLS
2000	0.6285	0.5499
	(0.0021)	(0.0015)
1999	0.5905	0.5118
	(0.0021)	(0.0015)
1998	0.5440	0.4706
	(0.0022)	(0.0015)
1997	0.4965	0.4272
	(0.0022)	(0.0015)
1996	0.4602	0.3944
	(0.0022)	(0.0016)
1995	0.4400	0.3743
	(0.0023)	(0.0016)
1994	0.4100	0.3519
	(0.0021)	(0.0015)
1993	0.3794	0.3386
	(0.0021)	(0.0015)
1992	0.3526	0.3140
	(0.0021)	(0.0015)
1991	0.3258	0.2903
	(0.0021)	(0.0015)
1990	0.2858	0.2547
	(0.0020)	(0.0015)
1989	0.2309	0.2073
	(0.0021)	(0.0015)
1988	0.1648	0.1443
	(0.0021)	(0.0015)
1987	0.1341	0.1158
	(0.0021)	(0.0015)
1986	0.1017	0.0857
	(0.0021)	(0.0015)
1985	0.0744	0.0624
	(0.0021)	(0.0015)
1984	0.0381	0.0355
	(0.0021)	(0.0015)
1983	.	.
M_UNION	.	0.1915
	.	(0.0009)
M_EXP	.	0.0290
	.	(0.0001)
M_EXP2	.	-0.0004
	.	(0.0000)
M_SCHOOLLING	.	0.0500
	.	(0.0001)
M_WHITE	.	0.0957
	.	(0.0010)

M_SMSASTAT	.	0.1101
	.	(0.0008)
M_FT	.	0.2055
	.	(0.0012)
F_UNION	.	0.2012
	.	(0.0011)
F_EXP	.	0.0189
	.	(0.0001)
F_EXP2	.	-0.0003
	.	(0.0000)
F_SCHOOLLING	.	0.0547
	.	(0.0001)
F_WHITE	.	0.0286
	.	(0.0010)
F_SMSASTAT	.	0.1303
	.	(0.0008)
F_FT	.	0.1093
		(0.0009)
R-squares	0.0929	0.5394

Table 3: Annual Wage Index Estimation Results

Variables	OLS With Only Year Dummies	OLS	CPI	ECI	GDP Deflator
2000	1.8749	1.7430	1.7289	1.8429	1.5366
1999	1.8048	1.6700	1.6727	1.7722	1.5139
1998	1.7229	1.6088	1.6365	1.7142	1.4945
1997	1.6429	1.5207	1.6114	1.6498	1.4756
1996	1.5844	1.4700	1.5753	1.5975	1.4510
1995	1.5528	1.4388	1.5301	1.5476	1.4235
1994	1.5069	1.4135	1.4880	1.5035	1.3941
1993	1.4614	1.3895	1.4508	1.4599	1.3627
1992	1.4228	1.3615	1.4086	1.4189	1.3291
1991	1.3852	1.3400	1.3675	1.3804	1.2990
1990	1.3308	1.3098	1.3122	1.3287	1.2528
1989	1.2597	1.2533	1.2450	1.2719	1.2069
1988	1.1792	1.1620	1.1878	1.2183	1.1619
1987	1.1435	1.1356	1.1406	1.1722	1.1257
1986	1.1071	1.1018	1.1004	1.1344	1.0966
1985	1.0772	1.0804	1.0803	1.0953	1.0711
1984	1.0388	1.0470	1.0432	1.0461	1.0711
1983	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4: Summary Statistics (N=2,940,849; Year 1983-2000)					
	Means				
Years	EARNWKE	AGE	EXP	EXP2	SCHOOLLING
2000	611.43	39.083	19.554	548.93	13.545
1999	585.81	38.986	19.455	543.33	13.546
1998	560.96	38.725	19.224	533.38	13.516
1997	527.67	38.506	19.031	523.88	13.490
1996	509.00	38.410	18.934	520.38	13.490
1995	498.36	38.393	18.907	518.02	13.500
1994	484.69	38.142	18.695	513.57	13.465
1993	469.74	37.949	18.569	508.96	13.398
1992	455.98	37.766	18.443	505.47	13.342
1991	443.76	37.564	18.309	503.63	13.281
1990	428.64	37.335	18.140	500.50	13.223
1989	406.09	37.088	17.926	495.84	13.192
1988	373.65	36.915	17.786	492.78	13.159
1987	362.64	36.727	17.629	488.04	13.130
1986	349.50	36.594	17.518	486.59	13.107
1985	340.10	36.573	17.545	491.99	13.060
1984	326.90	36.626	17.662	501.25	12.995
1983	311.29	36.812	17.885	514.84	12.959

Sources: Annual Earnings Files of the 1983 to 2000 Current Population Survey, Bureau of Census.

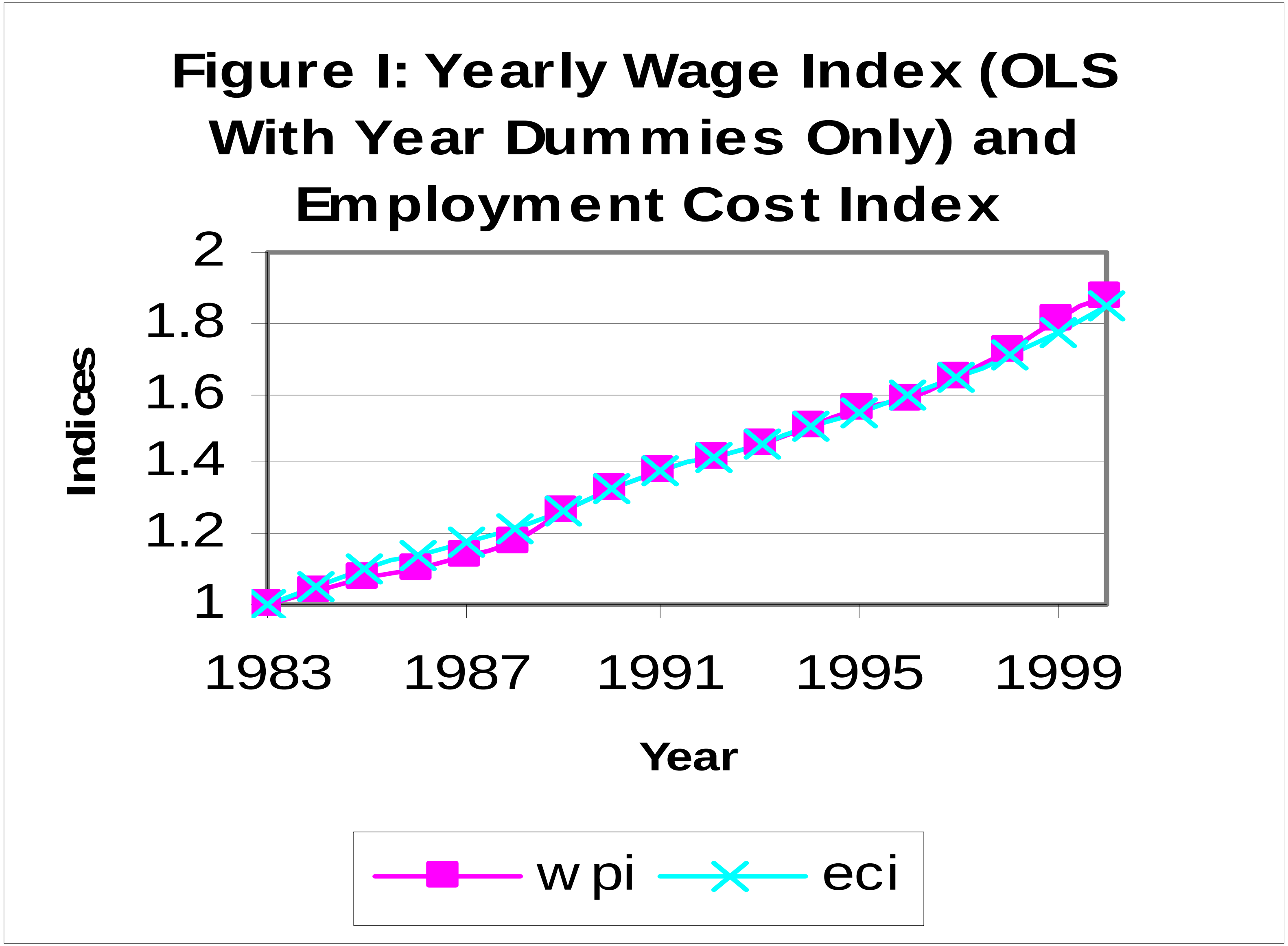
Figure I: Yearly Wage Index (OLS With Year Dummies Only) and Employment Cost Index
Indices
2
1.8
1.6
1.4
1.2
1
1983
1987
1991
1995
1999
Year
wpi
eci

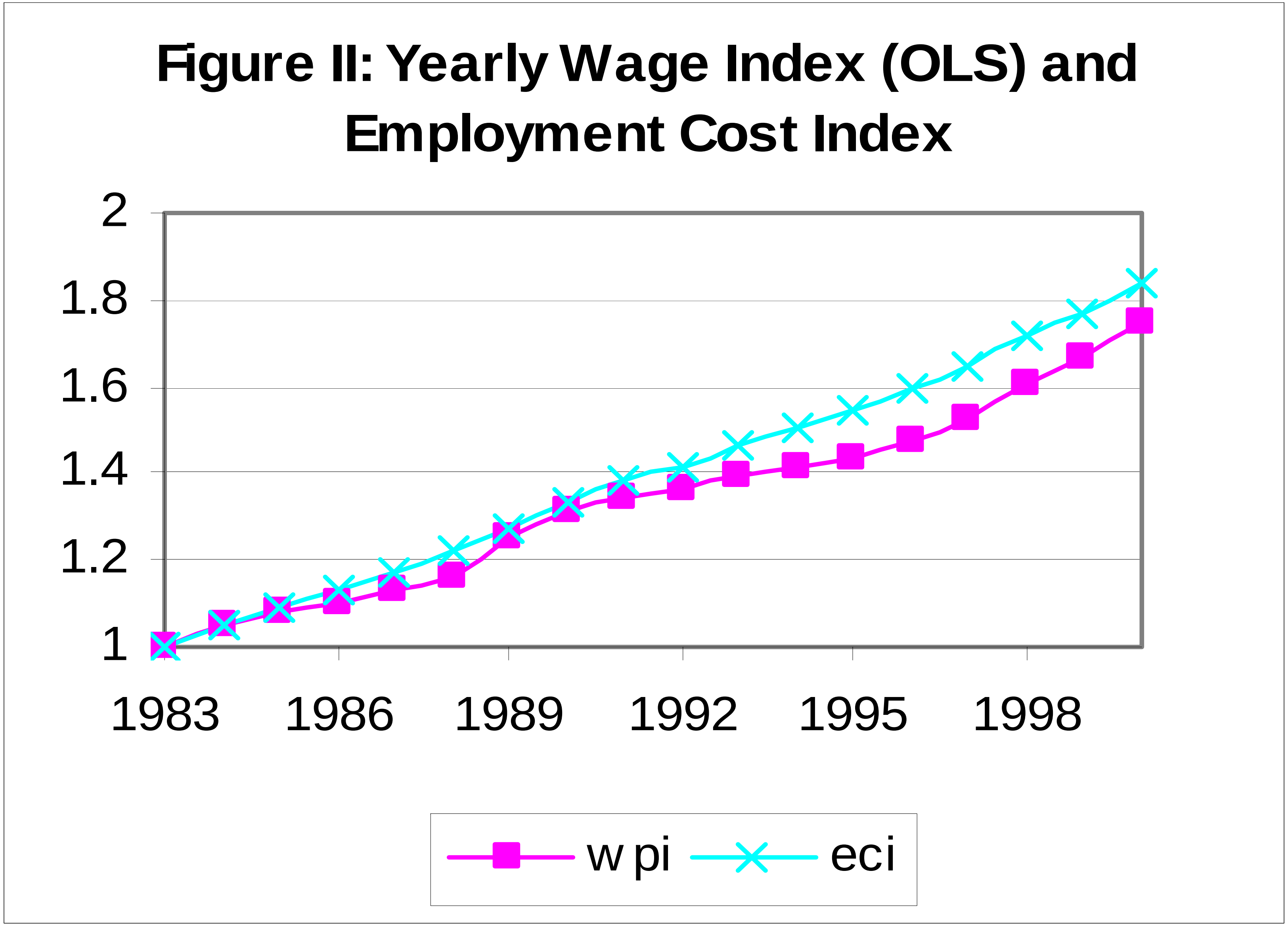

Figure II: Yearly Wage Index (OLS) and Employment Cost Index
2
1.8
1.6
1.4
1.2
1
1983
1986
1989
1992
1995
1998
wpi
eci