ESTIMATING UNEMPLOYED ENTRANTS INTO THE LABOR FORCE

By

Walter J. Sylva Jr.

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Masters of Arts

in

Economics

APPROVED:

R. Waud

(Department Head)

T. Lutton

R. Theroux

April 22, 2002

Fall Church, VA

Keywords: Entrants, Unemployment, Labor Force, Unemployment Insurance

ESTIMATING UNEMPLOYED ENTRANTS INTO THE LABOR FORCE

Walter J. Sylva Jr.

(Abstract)

This paper examines the methodology currently used by the Bureau of Labor Statistics to estimate unemployed entrants into the labor force. Unemployed entrants are individuals who enter the labor force for the first time, or after an absence of a year or more. Since these individuals lack current work experience, they are not eligible to receive unemployment compensation and therefore are not accounted for in the Unemployment Insurance system. According to the national Current Population Survey, unemployed entrants make up approximately 40 percent of the total unemployed. The estimation of unemployed entrants is a crucial step in the development of unemployment estimates at the labor market level. This paper demonstrates that the current method in use does not produce realistic estimates. Inconsistencies in the methodology are discussed and an econometric model is presented that produces more accurate measurements of the number of unemployed entrants into the labor force.

This paper is dedicated to my wife Christina
for all her support and encouragement.

Table of Contents

List of Tables

List of Charts

List of Equations

I. Introduction

The purpose of this paper is to demonstrate that a more accurate method exists to produce estimates of unemployed entrants into the labor force than the one currently used. Unemployed entrants are an important component in the estimation methodology for unemployment estimates and the unemployment rate at sub-state levels. The current methodology in use to estimate entrants dates back to the early 1980's and does not accurately reflect labor market trends.

Accuracy of the unemployment statistics is important at sub-state levels because they are one of the few economic indicators available for counties and cities. The Federal government uses these unemployment statistics to allocate funding for various entitlements and assistance programs. In fiscal year 2001, over $25 billion in federal funds were distributed based on these estimates.

Entrants are individuals who enter the labor force for the first time, or after an absence of a year or more. Individuals who are considered entrants include those who have recently completed military service, family responsibilities, education, and other reasons.[1] If they cannot find work, they are considered unemployed if they are able and willing to work and meet job search criteria as specified by state law. What makes these individuals different from other unemployed persons is that they often have had no previous work experience before their current spell of unemployment. If the individual had held a job for at least a year prior to becoming unemployed, they would most likely have qualified to receive unemployment compensation through their state and thus would be counted in the Unemployment Insurance (UI) system. Since entrants are not documented in the UI system, the counts of unemployed entrants into the labor force must be estimated each month.

[1] U.S. Department of Labor, Bureau of Labor Statistics, *Manual for Developing Local Area Unemployment Statistics*, December 1982, p 2-143

Before delving further into the current methodology used to estimate unemployed entrants in the labor force, some definitions must be discussed to familiarize the reader with the terminology and concepts used in labor force estimation. These terms and their definitions are standard concepts from the Current Population Survey (CPS). The CPS is a household survey that is the source for all national labor force data. The CPS is conducted monthly and reflects activities that occurred in the week that includes the 12th of the month (the reference week). All other labor force estimates produced for the state and sub-state levels use the same standard CPS concepts to be comparable to the CPS.

Unemployed refers to all persons who were not employed during the reference week but were available for work (excluding temporary illness) and had made specific efforts to find employment during the 4-week period ending with the reference week.[2]

Employed persons are those who, during the reference week (a) did any work at all (for at least 1 hour) as paid employees; worked in their own businesses, professions, or on their own farms; or worked 15 hours or more as unpaid workers in an enterprise operated by a family member or (b) were not working, but who had a job from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor management dispute, job training, or other family or personal reasons whether they were paid for the time off or were seeking other jobs. [3]

The *labor force* is the sum of employed persons and unemployed persons. The labor force consists of the non-institutional population, 16 years and older. Persons on active duty in the military are not included, nor are inmates of prisons, hospitals or residents of nursing homes.

For individuals without a job to be considered part of the labor force they must be able and willing to work, and be actively seeking employment. If they do not meet these

---

[2] U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, Economic and Statistics Administration, U.S. Census Bureau, Current Population Survey Design and Methodology, Technical Paper 63, 2000, p. 5-5.

[3] U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, Economic and Statistics Administration, U.S. Census Bureau, Current Population Survey Design and Methodology, Technical Paper 63, 2000, p. 5-3.

criteria, they are considered *not in the labor force*. Examples of individuals not in the labor force are persons not interested in working, such as retirees or full-time students. Another category in this group is discouraged workers. These are individuals who would like a job but have given up looking for work.

Unemployed entrants are broken down into two groups, new entrants and reentrants. New entrants are individuals entering the labor force for the first time. This group includes students who just have graduated high school or college and have never worked before. Reentrants are individuals who enter the labor force after a period of a year or more without working. Individuals in either of these groups are ineligible for UI benefits.

In accordance with state laws, each state determines if the individual is covered by the UI system and, if so, how much in benefits the individual is eligible to receive. Eligibility for benefits is based on separation and monetary issues.

Separation issues are concerned with the circumstances in which an individual lost their job. The UI system is designed to assist individuals who loose their job through no fault of their own. Those individuals who are fired for disciplinary or criminal reasons or if they voluntarily quit without good cause would not be eligible for compensation.

Monetary eligibility for benefits is determined by the amount of employment (in weeks or quarters) and wages earned by the individual (in some combination of dollars and time worked) in a specific base period.

A base period is a period of time in which a individual filing a claim must have had a specified minimum amount of insured (UI covered) work in order to qualify for benefits. Wages earned during this period are used in determining the claimant's weekly benefit amount and the claimant's maximum total annual benefits. In the majority of states, the base period is the first four quarters of the last five calendar quarters.

Based on national data from the Current Population Survey (CPS) the unemployed entrants into the labor force represent about 40 percent of total unemployment (*See Table 1*.). The CPS is the only reliable published source of national data on new entrants and reentrants into the labor force. Entrants data from the CPS for

states are not published because these estimates are not statistically reliable on a monthly basis. However, reliable estimates of entrants do exist but are limited to only the four most populous states. Entrants data from the CPS do not exist for labor market areas (LMA).

Table 1. Entrants as a percent of total unemployed, United States, annual averages, 1980-2000

Year	CPS Unemployed *(in thousands)*	CPS Entrants *(in thousands)*	Entrants as a percent of total unemployed
1980	7637	2799	36.7
1981	8273	3083	37.3
1982	10678	3569	33.4
1983	10717	3628	33.9
1984	8539	3294	38.6
1985	8312	3295	39.6
1986	8237	3189	38.7
1987	7425	2894	39.0
1988	6701	2625	39.2
1989	6528	2520	38.6
1990	7047	2618	37.2
1991	8628	2931	34.0
1992	9613	3222	33.5
1993	8940	3117	34.9
1994	7996	3390	42.4
1995	7404	3104	41.9
1996	7236	3092	42.7
1997	6739	2907	43.1
1998	6210	2652	42.7
1999	5880	2474	42.1
2000	5655	2388	42.2

Source: Bureau of Labor Statistics

The estimation of unemployed entrants into the labor force is a crucial step in determining the level of unemployment and the unemployment rate for the regional labor markets. The quality of these estimates is very important because they are the only economic indicators for these geographic areas and are the major inputs into funding algorithms used to allocate federal funds to numerous assistance and entitlement programs.

II. Estimation Background

Each month the Bureau of Labor Statistics (BLS) produces estimates of the number of unemployed persons for three geographic levels. These geographic levels consist of the nation; each of the 50 states (including the District of Columbia and Puerto Rico), and sub-states areas comprised of labor market areas; and their component counties and cities.

A hierarchy of estimating procedures exists for each of these geographical levels starting with the use of the CPS for the national labor force estimates, followed by a time series econometric model for state estimates, and finally the application of what is called the handbook method for sub-state estimates.

Monthly estimates for the nation as a whole are derived each month from the CPS. The CPS is a monthly survey of approximately 60,000 households. The reference period covered by the survey is the week that includes the 12th of the month. More detailed documentation on the CPS can be found in the Chapter 1 of the BLS Handbook of Methods published by the U.S. Department of Labor.

Estimates for individual states are not derived from the CPS. The CPS sample sizes for individual states are not large enough to produce statistically reliable monthly estimates. (The exceptions are the four most populous states, California, Florida, New York, and Texas. However, these data are not the official estimates and are not published.) The official monthly labor force estimates for all states are produced using a variable coefficient time-series regression model. This model utilizes input data from Unemployment Insurance (UI) claim counts and employment data from the Current Employment Statistics (CES) program of the BLS to predict the monthly CPS labor force estimates for each state.

Labor market area estimates are created each month using local data from various sources including the UI and CES programs. This methodology is referred to as the handbook method. The handbook method consists of sixteen steps that calculate

employment and unemployment estimates separately and then combine these estimates to create the labor force and unemployment rate for a labor market area.

Information on the state models and the handbook methodology is documented in numerous U.S. Department of Labor publications including, *How the Government Measures Unemployment*, *Estimating Unemployment*, and Chapter 4 of the BLS Handbook of Methods.

This paper examines one component of the unemployment estimation methodology, that is the estimation of unemployed entrants into the labor force. However, an understanding of entire handbook estimation process is needed for the reader to appreciate how this one component fits into the process and how it affects the labor force estimates at the sub-state level. Therefore, a brief description of the handbook methodology is in order.

The total number of labor market areas in each state exhausts the entire geography of the state. The number of employed and unemployed are estimated for each labor market area. These estimates are then used to derive the labor force level and the unemployment rate for the labor market area.

To arrive at a total employment level for a labor market area, payroll employment from the CES is combined with an agricultural employment estimate and an estimation of *all other* employment. The agricultural employment estimate is derived from the U.S. Department of Agriculture's annual Agricultural Labor Survey (ALS). *All other* employment includes workers not covered in the CES or the ALS, such as the self-employed, unpaid family workers, and domestic workers in private households.

The estimate of unemployment is comprised of UI continued claims data, UI final payments, and unemployed entrants into the labor force.

Continued claimants are persons certifying to a week of unemployment. Only continued claims for the week of the 12th are used to comply with the CPS reference period. Continued claims data come from three separate UI programs. These programs are the regular state UI system, Unemployment Compensation for Federal Employees (UCFE), and claims data from the Railroad Retirement Board. (The Railroad Retirement

Board is responsible for administrating unemployment compensation for railroad workers.)

Final payments are identified when claimants receive their final benefit payment through the UI system. Each month the final payments data are added to a pool of previous final payments that are carried forward using a survival rate that reflects the typical duration of unemployment after benefits have been exhausted.

The most difficult component of the handbook method is the estimation of unemployed entrants into the labor force. Entrants are individuals with no previous work experience or who have not held a job in over a year. Since they have not worked, they have not contributed to the UI fund through their payroll taxes. Entrants are therefore ineligible for UI benefits and are not tracked in the UI system. (Details of the procedures used to estimate unemployed entrants are discuss in *Section III* of this paper.)

Once the employment and unemployment levels have been compiled, the labor force level and unemployment rate can be derived for the labor market area. This estimate is commonly referred to as the handbook estimate because the methodology is based on procedures that originated from a 1960's publication from the U.S. Department of Labor entitled *Handbook on Estimating Unemployment*. It is also called an independent estimate since it is derived independently of the state estimates.

After the handbook estimates are calculated for each labor market area in a state, they are adjusted to match the state totals produced from the time-series models. This procedure is known as additivity because it forces the independent labor market area estimates to add the statewide estimates. To obtain a labor force estimate for a given area, a handbook share is computed for that area. A handbook share is the ratio of the area's handbook estimate to the sum of the handbook estimates for all labor market areas in the state. This ratio is then applied to the current statewide estimate to produce the final labor market area estimates (*See Table 2*.).

"The Handbook-Share method of apportioning the State estimates of unemployment and employment to areas assumes a proportional distribution throughout the State of the difference between the sum of sub-state Handbook estimates and the

independent State estimates. This adjustment is performed for both preliminary and revised estimates."[4]

Table 2. Handbook-share method

	Unemployment			Employment		
Area	***Handbook***	***Percent of Summed Handbook***	***Statewide****	***Handbook***	***Percent of Summed Handbook***	***Statewide****
State			*49,300*			*562,800*
MSA 1	18,500	0.394456	19,447	190,600	0.3481279	195,926
Major LMA 1	9,300	0.198294	9,776	107,100	0.1956164	110,093
Major LMA 2	8,700	0.185501	9,145	103,400	0.1888585	106,290
Minor LMA 1	2,300	0.049041	2,418	36,800	0.0672146	37,828
Minor LMA 2	1,900	0.045120	1,997	25,900	0.0493059	26,624
Intrastate Portion of Interstate MSA 2	6,200	0.132196	6,517	83,700	0.1528767	86,039
Sum of substate Areas	46,900	1.000000	49,300	547,500	1.000000	562,800

*For the State, enter the model-based estimate. The substate data are the product of the area's Percent of Summed Handbook and the statewide estimate for unemployment or employment.

Source: Local Area Unemployment Statistics Program Manual

Once the additivity process is completed, labor force estimates for cities and towns with populations of 25,000 or more are disaggregated from their labor market areas. The disaggregation process breaks out the number of employed and unemployed based on the component areas' relationship to the labor market area. Two methods are used: a population-claims method or a census-shared method. The population-claims method is the preferred procedure since it uses current UI claims and population data. However, if claims data are not available at the city and town level then the census-shared method is used. This method is less preferred, since it uses information based on the component areas' relationship to the labor market area at the time the census was conducted. For counties and cities, these estimates are the only indicators available that can identify the economic wellbeing of a local area.

[4] U.S. Department of Labor, Bureau of Labor Statistics, *Local Area Unemployment Statistics Program Manual*, 2001, p. 8-3

III. Overview of Current Methodology

New entrants and reentrants into the labor force cannot be derived from UI statistics, because these individuals are ineligible for UI compensation and are not entered into the UI system. Instead, entrants into the labor force are estimated on the basis of the national historical relationship of entrants to the experienced unemployed and the experienced labor force. The current methodology in use to estimate entrants into the labor force is a function of the particular month of the year, the level of the experienced unemployed, the level of the experienced labor force, and the youth proportion of the working-age population[5]. This last component is known as the YPR and is defined as the population of the 16-19 age group divided by the population of the 20 and over age group.

The estimate of total entrants for a given month is derived from *equation 1*.[6]

$$\mathbf{ENT = A(X+E)+BX} \qquad \text{(EQ 1.)}$$

Where:

ENT	=	total entrant unemployment
E	=	total employment
X	=	total experienced unemployment
A,B	=	synthetic factors incorporating both seasonal variations and the assumed relationship between the proportion of youth in the working age population and the historical relationship of entrants to either the experienced unemployed (B factor) or the experienced labor force (A factor)

The A and B factors of *equation 1* are the products of two separate linear regression models that are used to develop annual factors. These annual factors are then seasonally adjusted and applied to the experienced unemployed and the experienced labor force. The output of these models is used in the estimation of unemployed in the handbook procedure for labor market areas.

---

[5] U.S. Department of Labor, Bureau of Labor Statistics, BLS Handbook of Methods, Bulletin 2490, 1997, p. 39

The A factor represents the relationship of entrants to the labor force and is expressed in *equation 2*. The fit of the A factor model is $R^2 = 0.49$, which was derived from the model's published correlation coefficient of 0.70.[7]

A factor = -0.019885 + 0.011151 ln YPR (EQ 2.)

Once computed, the A factor is multiplied by a seasonal factor derived from the CPS referred to as A′ and then applied to the "experienced" labor force.

A factor (A′) experienced labor force =
the relationship of entrants to the experienced labor force (EQ 3.)

The experienced labor force is the sum of the total employed and the experienced unemployed. The experienced unemployed is simply the number of individuals who are receiving UI compensation (continued claims).

The B factor represents the relationship of entrants to the experienced unemployed and is expressed in *equation 4*. The B factor model has a fit of $R^2 = 0.7225$. This was derived from the B factor model's correlation coefficient of 0.85.[8]

B factor = -0.3987 + 0.2271 ln YPR (EQ4.)

The B factor is multiplied by a seasonal B′ factor and then applied to the experienced unemployed or the number of individuals who are receiving continued claims.

B factor (B′) experienced unemployed =
the relationship of entrants to the experienced unemployed (EQ 5.)

The estimated A factor and B factor levels of unemployment are added together to arrive at the number of unemployed entrants into the labor force. The estimate of entrants is then added to the monthly UI claims data and the final payment estimates to arrive at the monthly unemployment level for a labor market area.

---

[6] U.S. Department of Labor, Bureau of Labor Statistics, BLS Handbook of Methods, Bulletin 2490, 1997, p. 39

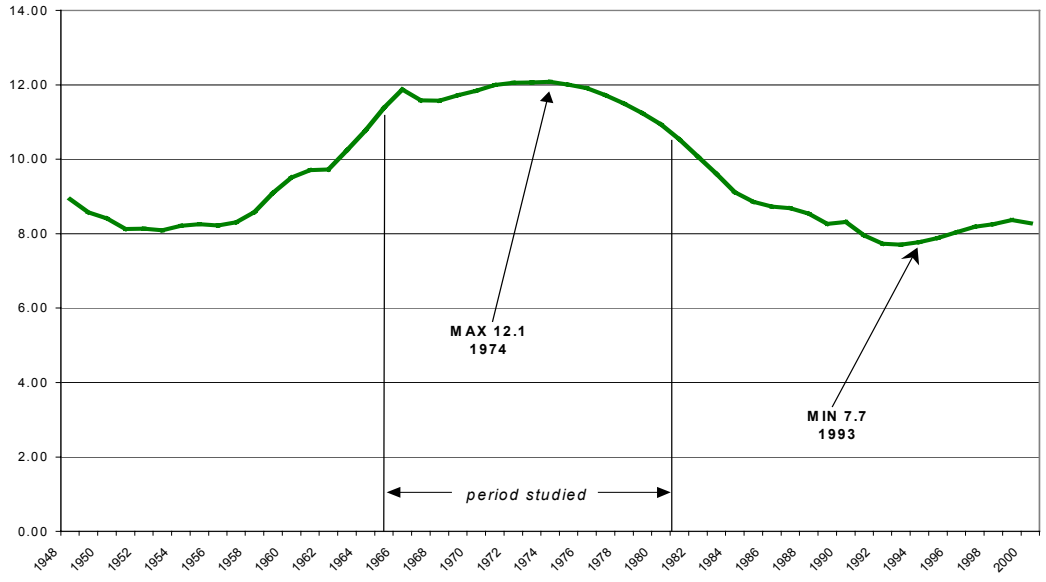
[7] U.S. Department of Labor, Bureau of Labor Statistics, *Manual for Developing Local Area Unemployment Statistics*, December 1982, p 2-151.2

III. Inconsistencies with Current Methodology

The A and B factor models rely solely on the youth population ratio (YPR) as the only information to explain any movement in new and reentrants. These models were developed in the early 1980's based on a study of national data from 1965 to 1981 and do not accurately reflect the labor market situation.

During the period studied, the YPR was at an all time high ranging from 10.5 percent to 12.1 percent. From 1981 to 2000, the YPR has ranged from 7.7 percent to 8.9 percent, more in line with the period prior to 1965. *(See Chart 1.)*

Chart 1. Annual Youth Population Ratios*, United States, 1948-2000



**Derived from the CPS* *Source: Bureau of Labor Statistics*

[8] U.S. Department of Labor, Bureau of Labor Statistics, *Manual for Developing Local Area Unemployment Statistics*, December 1982, p 2-151.2

The Manual for Developing Local Area Unemployment Statistics states that "the labor force participation of workers marginally attached to the labor force is affected by cyclical changes in the economy."[1] However, the A and B factors produced by the current models do not exhibit cyclical behavior.

To illustrate this point, the A and B factors were revised using the actual annual CPS data from 1967 to 2000 rather than the output from the respective regression models.

The A factors were updated by dividing CPS unemployed entrants by the experienced labor force. Subtracting the CPS new entrants and the CPS re-entrants from the CPS total labor force produced the experienced labor force data.

The B factor was updated by dividing CPS unemployed entrants by the experienced unemployed. Subtracting the CPS new entrants and the CPS re-entrants from the CPS total unemployment produced the experienced unemployed numbers.

The new factors resulting from the use of the actual CPS data are much larger than the factors currently in use and have more variation. Peaks and troughs in the updated factors tend to follow the business cycle of the nation's economy while the original factors remained relatively flat.

When the economy is expanding, the proportion of unemployed entrants into the labor force, the A factor, is low, indicating that the number of unemployed entrants entering the labor force has declined. In other words, most entrants into the labor force are able to find jobs in the growing economy.

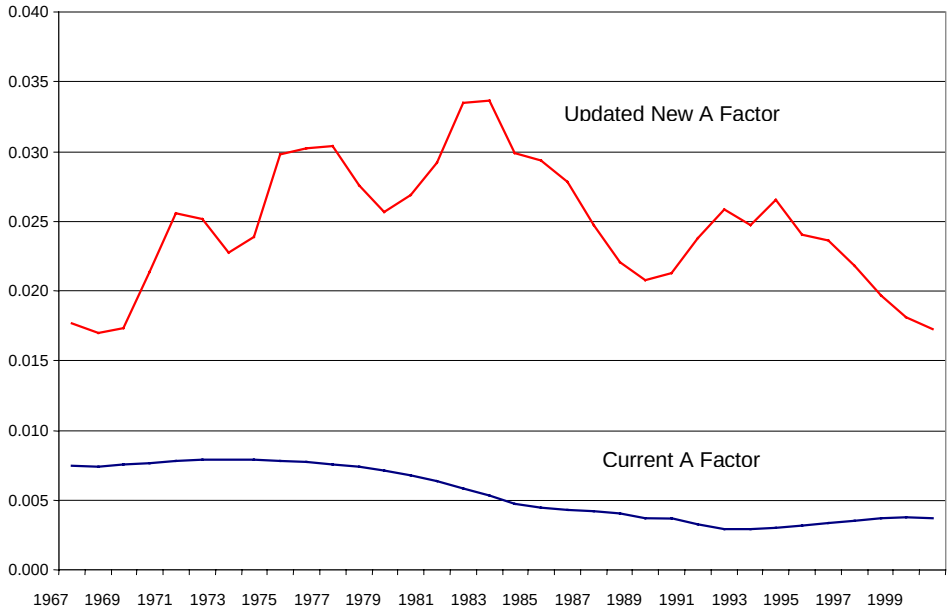
At the same time the number of unemployed entrants in relation to the experienced unemployed, the B factor, is large because the pool of experienced unemployed is smaller since less of the experienced workers are without jobs. An example of this is can be seen on *Charts 2 and 3* in December 1969 when the business cycle was at a peak. *(See Table 3. for business cycles.)*

---

[1] U.S. Department of Labor, Bureau of Labor Statistics, *Manual for Developing Local Area Unemployment Statistics*, December 1982, p 2-147

The opposite is true when the economy is contracting. In economic down-turns more entrants into the labor force are not able to find jobs. The A factor will increase as entrants cannot be absorbed into the experience labor force.

Chart 2. Comparison of annual A factors, 1967-2000



During a recession the B factor becomes smaller as the proportion of unemployed entrants to the experienced unemployed declines. This occurs because the number of unemployed entrants tends to fall, as young workers become discouraged from entering the labor force due to lack of job opportunities. Entrants without any work experience would have to compete with a larger pool of experienced unemployed who are now looking for jobs as well. This situation is known as the discouraged worker effect. An example of this behavior can be seen on the *Charts 2 and 3* in November 1982 which was a low point of the business cycle. (*See Table 3. for business cycles*.)

Chart 3. Comparison of annual B factors, 1967-2000

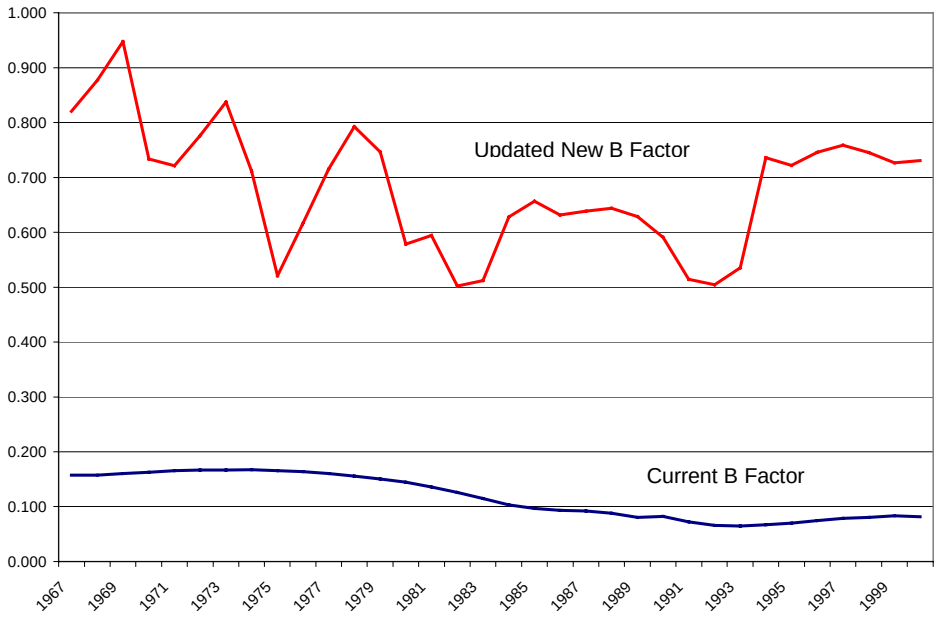


Table 3. Business cycle turning points, 1960-99

Peaks	Troughs
April 1960	
	February 1961
December 1969	
	November 1970
November 1973	
	March 1975
January 1980	
	July 1980
July 1981	
	November 1982
July 1990	
	March 1991

Source: Bureau of Economic Analysis

Another indication of the current methodological shortcomings can be seen in the additivity ratios for each state. The additivity ratios represent the share of the individual labor market area estimates developed from the independent handbook estimation to the state total produced by the time-series model. There is an employment additivity ratio and an unemployment additivity ratio for each labor market area in a state and for the

state as a whole. These ratios are used to force the sum of the labor market areas to equal the state totals from the time-series model.

In general, the employment additivity ratios for a given state tend to be close to 1. Thus when the independent labor market area employment estimates are summed, the sum closely matches the state employment estimates from the time-series model.

The additivity ratios for unemployment, on the other hand, tend to range between 0.99 and 3.8 depending on the state. For January 2001 for example, additivity ratios were produced for 335 labor market areas. *(See Table 4.)* Out of this total, 328 additivity ratios for unemployment had values greater than 1.1 and eight were greater than 3.0. This is a good indication that the handbook unemployment methodology is not accurately measuring the number of unemployed.

Table 4. Additivity Ratios for Selected Areas, January 2001, Ranked by Unemployment Ratio

Labor Market Area	Employment Ratio	Unemployment Ratio
New Orleans, LA MSA	0.969805	3.8280433
Yuma, AZ MSA	0.957450	2.81105121
Pensacola, FL MSA	0.978033	2.69201878
Manchester, NH PMSA	0.992402	2.54223149
Baltimore, MD PMSA	1.005159	2.3706741
Dallas, TX PMSA	0.993960	2.20515104
Los Angeles-Long Beach, CA PMSA	1.122231	2.20127248
Denver, CO PMSA	0.932755	1.82485199

Source: Bureau of Labor Statistics

I. Data Availability

It is clear that the YPR alone can not adequately explain the level of unemployed entrants into the labor force. Other data exist that can provide additional information to predict the number of entrants more accurately than the current models. The following data types were selected for testing as possible explanatory variables.

Population data were used since, as the previous study showed, there is a relationship between the level of unemployed entrants and the population. The civilian non-institutional population of persons 16 years and older was used to keep the estimates consistent with the CPS concepts of population and labor force. "In the CPS, labor force data are restricted to persons 16 years of age and older, who currently reside in one of the fifty states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces."[2]

This population group of the civilian non-institutional population of persons 16 years and older constitutes all persons of working age in the nation who could possibly be participating in the labor force. This population group is broken down into two categories similar to the YPR concept. The first category is the population of 16-19 year olds. This category represents teenagers who are now of legal working age and can enter the labor force. The second category is the population of persons who are age 20 years and older. This second category represents the experienced workers, those either employed or those who have held jobs.

UI claims data provide information on the number of individuals who file for benefits for the first time and the number whom are receiving benefits for certified weeks of unemployment.

[2] U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, Economic and Statistics Administration, U.S. Census Bureau, Current Population Survey Design and Methodology, Technical Paper 63, 2000, p. 5-3.

There are two types of claims data available, initial claims counts and continued claims counts. An initial claims is any notice of unemployment filed to request a determination of and eligibility for compensation. The weekly initial claims count is considered to be an economic indicator since it measures individuals who have recently lost heir jobs or who cant find work. A continued claim is a claim filed for benefit payment for one or more weeks of unemployment. Both claim types are published weekly in the US Department of Labor's Employment and Training Administration's (ETA) news release.

The unemployment rate, another major economic indicator, also was considered to provide additional information in explaining entrants. Since the unemployed entrants into the labor force is actually a component of the total unemployed, the current month's unemployment rate could not be used, so a one-month lagged unemployment rate was used. It is also possible that the previous published unemployment rate could have an effect on an individual's decision to enter the labor force.

VI. Entrants / Population Model

The entrants/population model incorporates all the data elements described in the previous section into a single model to predict the unemployed entrants into the labor force. The model was tested using data at the national level.

Since it is the goal of the author to create a model that will be applicable to predicting entrants on the state level and on the labor market level, ratios were used rather than the actual level data. These ratios are based on the variable's relationship to the experienced worker's population group, or the 20 years and older population. If the actual levels were used, the model would yield parameters that would be too large to apply to data for any geographic areas smaller than the nation. To obtain the level estimates, the dependent variable simply needs to be multiplied by the area's population of 20 years and older for the corresponding month.

In addition to transforming the explanatory variables into ratios, another adjustment was made to the explanatory variables. In order to avoid any suspicion of reverse causation, each variable was lagged by one month. In other words, the number of entrants into the labor force may have an affect on the number of initial claim, continued claims, or the unemployment rate within the same reference period.

Seasonal fluctuations in the data were accounted for by introducing dummy variables into the model. Seasonal fluctuations cause a regular rise and fall in the data series that recurs each year but are not more than a one-year duration. These fluctuations are due to the influences of weather, holidays, the opening and closing of schools, and other recurring seasonal events. To identify seasonal movements, a dummy variable was created for each of the first 11 months of the year, January through November. (December is implicitly represented as the month when all the dummy variables are equal to zero.) When the observation for a particular month matches the corresponding monthly dummy variable, a value of 1 is associated with that month, while the remaining months retain a zero value.

The complete model is specified as follows:

$$\mathbf{ENTR/Pop_{20+}} = \beta_0 + \beta_1(\mathbf{IC/Pop_{20+}})_{t-1} + \beta_2(\mathbf{CC/Pop_{20+}})_{t-1} + \beta_3(\mathbf{Pop_{16-19}/Pop_{20+}})_{t-1} + \beta_4(\mathbf{UR_{t-1}}) + \beta_5(\mathbf{D_1}) + \beta_6(\mathbf{D_2}) + \beta_7(\mathbf{D_3}) + \beta_8(\mathbf{D_4}) + \beta_9(\mathbf{D_5}) + \beta_{10}(\mathbf{D_6}) + \beta_{11}(\mathbf{D_7}) + \beta_{12}(\mathbf{D_8}) + \beta_{13}(\mathbf{D_9}) + \beta_{14}(\mathbf{D_{10}}) + \beta_{15}(\mathbf{D_{11}}) + \varepsilon \qquad \text{(EQ 6.)}$$

where: $\mathbf{ENTR/Pop_{20+}}$ = new and re-entrants into the labor force from the CPS divided by CPS population of 20 years and older.

$(\mathbf{IC/Pop_{20+}})_{t-1}$ = initial claims divided by CPS population of 20 years and older, lagged one month.

$(\mathbf{CC/Pop_{20+}})_{t-1}$ = continued claims divided by CPS population of 20 years and older, lagged one month.

$(\mathbf{Pop_{16-19}/ Pop_{20+}})_{t-1}$ = CPS population of 16-19 years old divided by CPS population of 20 years and older, lagged one month.

$\mathbf{UR_{t-1}}$ = CPS unemployment rate, one month lagged

$\mathbf{D_1}$ = seasonal dummy variable for January

$\mathbf{D_2}$ = seasonal dummy variable for February

$\mathbf{D_3}$ = seasonal dummy variable for March

$\mathbf{D_4}$ = seasonal dummy variable for April

$\mathbf{D_5}$ = seasonal dummy variable for May

$\mathbf{D_6}$ = seasonal dummy variable for June

$\mathbf{D_7}$ = seasonal dummy variable for July

$\mathbf{D_8}$ = seasonal dummy variable for August

$\mathbf{D_9}$ = seasonal dummy variable for September

$\mathbf{D_{10}}$ = seasonal dummy variable for October

$\mathbf{D_{11}}$ = seasonal dummy variable for November

The model was estimated using the Ordinary Least Squares (OLS) regression method for monthly observations from January 1980 to December 2000. The results were a model fit of $R^2 = 0.9018$ with most of the variables having strong t-values, with the exception of $(IC/Pop_{20+})_{t-1}$. The variable parameters are listed below in *Table 5*. Although these results looked promising, testing for autocorrelation using the Durbin-Watson test revealed that first order autocorrelation exists.

To correct for first order autocorrelation, the Generalized Least Squares (GLS) regression method was used. The GLS method removed much of the serial correlation. The *d* statistic of the Durbin-Watson test was 2.26 compared to 0.741 using the OLS

method. As consequence of using GLS, the R^2 was slightly lowered to 0.8214. (*See Table 6.*)

Table 5. Entrants/Population Model Parameters, ***Ordinary Least Squares***.

Variable	Parameter Estimate	Standard Error	t Value
Intercept	0.001368	0.000768	1.78
$(IC/Pop_{20+})_{t-1}$	0.1556	0.2942	0.53
$(CC/Pop_{20+})_{t-1}$	-0.2157	0.0536	-4.02
$(Pop_{16-19}/Pop_{20+})_{t-1}$	0.0403	0.0106	3.8
UR_{t-1}	0.00209	9.48E-05	22.05
D_1	0.002552	0.000347	7.36
D_2	0.001101	0.000454	2.42
D_3	0.001182	0.000382	3.1
D_4	0.000813	0.000397	2.05
D_5	0.00309	0.000357	8.67
D_6	0.006247	0.000346	18.05
D_7	0.003555	0.000339	10.5
D_8	0.002144	0.000349	6.15
D_9	0.002298	0.000344	6.67
D_{10}	0.001847	0.000341	5.41
D_{11}	0.001369	0.000336	4.07

Table 6. Comparison of Regression Methods

Regression Method	Durbin-Watson Test (*d*)	Fit (R^2)
OLS	0.741	0.9018
GLS	2.23	0.8214

Using the GLS regression method, the parameter estimates changed along with their t-values. (*See Table 7.*) The t-values actually increased for nine of the sixteen variables, while the rest declined. Of the variables that experienced a decrease in their t-values, three variables decreased in value to where they were no longer significant. These were $(CC/Pop_{20+})_{t-1}$, D_3 and D_4.

Table 7. Entrants/Population Model Parameters,
Generalized Least Squares.

Variable	Parameter Estimate	Standard Error	t Value
Intercept	0.001853	0.001404	1.32
$(IC/Pop_{20+})_{t-1}$	-0.3612	0.2487	-1.45
$(CC/Pop_{20+})_{t-1}$	-0.00988	0.0662	-0.15
$(Pop_{16-19}/Pop_{20+})_{t-1}$	0.0452	0.0189	2.39
UR_{t-1}	0.001644	0.000155	10.61
D_1	0.002363	0.000242	9.76
D_2	0.001231	0.000425	2.90
D_3	0.000608	0.000366	1.66
D_4	0.000136	0.00038	0.36
D_5	0.002623	0.000336	7.81
D_6	0.006021	0.000315	19.11
D_7	0.003648	0.000314	11.62
D_8	0.002288	0.000312	7.33
D_9	0.002153	0.00029	7.42
D_{10}	0.001881	0.000267	7.04
D_{11}	0.001398	0.000208	6.72

Examining each of the coefficients for the variables reveals that they are consistent with a priori expectations.

The initial claims coefficient shows a negative relationship to the dependent variable, indicating that as initial claims increase the number of entrants decrease. However, the t-statistic is only significant at the 90 percent level. Initial claims are an unemployed individual's first request for UI compensation and also a major economic indicator that is published weekly. An increase in initial claims indicates deteriorating employment opportunities which tend to cause individuals who are marginally attached to the labor force not to participate because of poor job prospects. For example, students may choose to stay in school rather than enter a weak job market.

Continued claims have a negative relationship with the dependent variable. Continued claims represent the experienced unemployed, which are those unemployed individuals with prior work experience. When the number of continued claims rises, as in a recession, more individuals with work experience are without jobs and are collecting UI compensation. As continued claims counts rise, entrants into the labor force tend to

decline as they become discouraged by the limited job prospects which are filled by those with prior work experience. However, the t-statistic indicates that this variable is not significant.

The population variable, $(Pop_{16\text{-}19}/Pop_{20+})_{t\text{-}1}$, has a positive relationship with the dependent variable. As the 16 to 19 year old population group increases, the number of unemployed entrants into the labor force increases. This age group represents teenagers who are now of legal working age and can enter the labor force, but are not finding jobs. This variable is also defined as the youth population ration that was the sole explanatory variable in the original A and B factor models described in Section III.

There is a positive relationship for the lagged unemployment rate. As the unemployment rate increases, so does the number of unemployed entrants into the labor force. This is a logical relationship since the unemployment rate comes from the CPS and, as was previously noted, unemployed entrants make up approximately 40 percent of the CPS unemployment.

Looking at the seasonal dummy variables, we see larger coefficients for months that have large seasonal movement in the data. The number of unemployed entrants tends to peak each year in certain months. This especially noticeable in the summer months entrants when students enter the job market. *Chart 4* plots the dummy variable coefficients and for the most part illustrates the seasonal patterns that occur in the number of unemployed entrants into the labor force. *Chart 5* shows the CPS unemployed entrants for the period of 1980 through 2000 averaged by month. The averaged entrants data in chart 5 displays a similar pattern to the dummy variables. The exception is the month of January. However, the higher coefficient value for D_1, which corresponds to January in *Chart 4*, can be explained by initial claims variable. The initial claims counts in January are typically large as the number of claims peak because workers who were hired for the holiday season are no longer needed.

Chart 4. Coefficients of the seasonal dummy variables

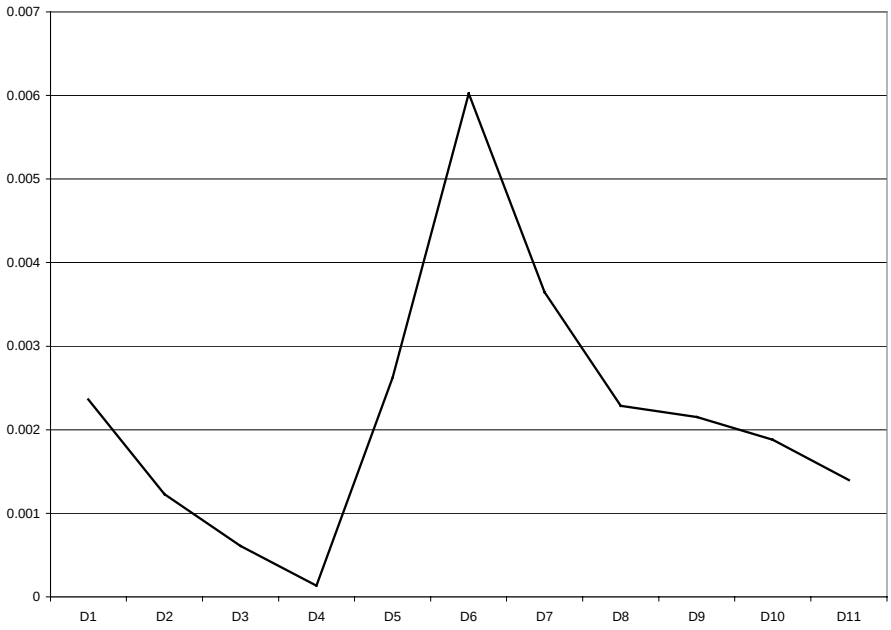


Chart 5. CPS entrants, 1980-2000 averaged by month

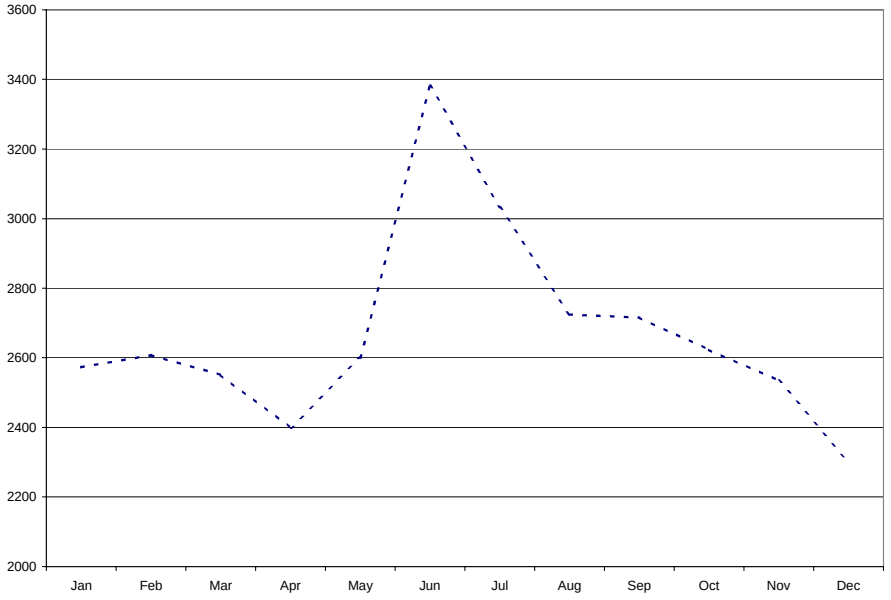
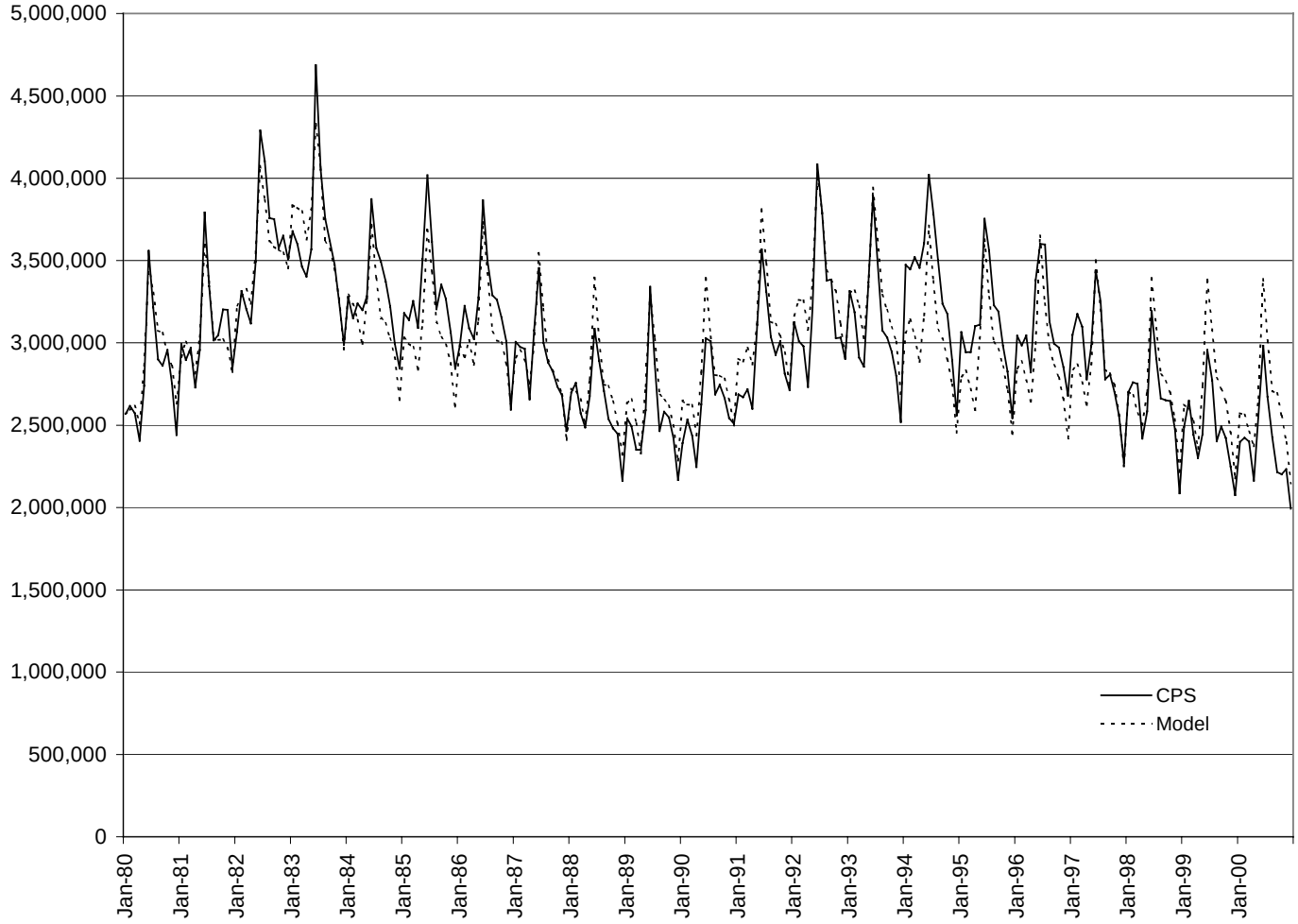


Chart 6 shows the in-sample model predictions produced from the parameters in *Table 7* versus the actual CPS level of entrants; the model estimates follow the CPS quite well.

Chart 6. Entrants, CPS vs Entrants/Population Model, United States, January 1980-December 2000



VII. Model Testing

The newly specified entrants/population model was first tested by forecasting 60 out-of-sample periods as described below. Next a rolling regression technique was applied to the model to further enhance the quality of the estimates. Due to the complexity of this technique, only 24 out-of-sample periods were used. Details on this procedure are described subsequent to those of the first test.

Forecasts for the period of January 1996 through December 2000 were generated for the newly specified model using national data from January 1980 through December 1995. The forecasted estimates were compared to the current method's estimates that were calculated for the same time period using equations 2 through 5 that were described in Section III of this paper. This out-of-sample period amounted to a total of 60 monthly observations. The predicted values of the entrants/population ratio were multiplied by the population of 20 years and older and were compared to the actual data from the CPS for the same time period. *Table 8* lists the model parameters used in the January 1996 through December 2000 forecasts.

Table 8. Entrants/Population Model Parameters, Generalized Least Squares, January 1980-December 1995.

Variable	Parameter Estimate	Standard Error	t Value
Intercept	0.001761	0.001497	1.18
$(IC/Pop_{20+})_{t-1}$	-0.3496	0.2877	-1.22
$(CC/Pop_{20+})_{t-1}$	-0.0525	0.0703	-0.75
$(Pop_{16-19}/Pop_{20+})_{t-1}$	0.0591	0.0199	2.97
UR_{t-1}	0.001573	0.000183	8.60
D_1	0.002426	0.000285	8.51
D_2	0.001369	0.000517	2.65
D_3	0.000644	0.000415	1.55
D_4	0.000343	0.000437	0.78
D_5	0.002713	0.000389	6.97
D_6	0.006406	0.000367	17.46
D_7	0.003796	0.000366	10.37
D_8	0.002565	0.000367	6.99
D_9	0.002372	0.000337	7.04
D_{10}	0.001969	0.000309	6.37
D_{11}	0.001401	0.000243	5.77

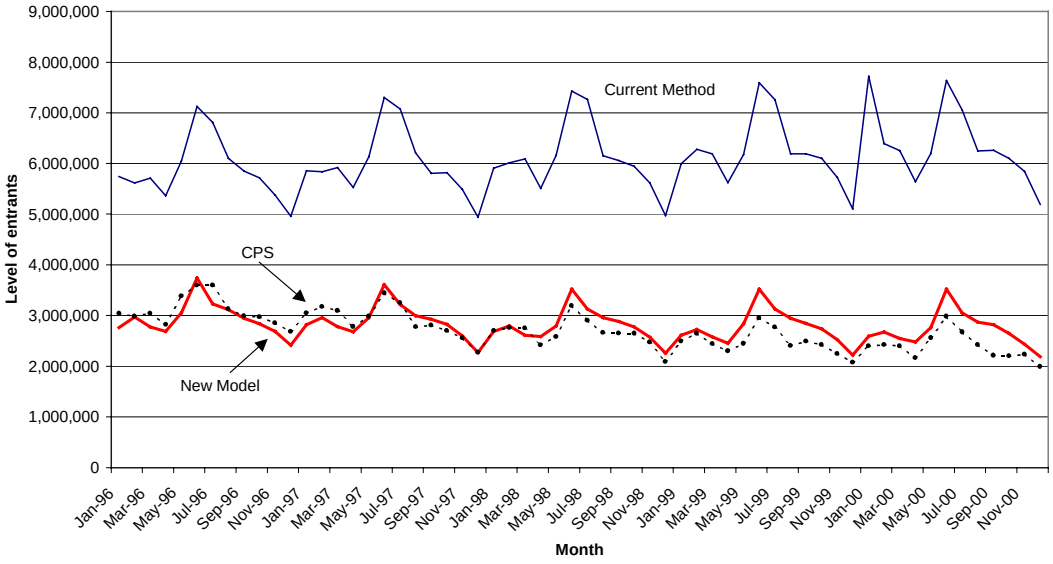
Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were calculated for both methods and are listed in *Table 9.* Both measurements of forecasting error were significantly smaller for the new model compared to the current model, indicating that the entrants/population model produces much more accurate prediction of unemployed entrants into the labor force than the current method.

Table 9. Forecasting Errors, National Data, January 1996-December 2000

Methodology	RMSE	MAE
Current Method	3,458	3,404
New Model	262	219

In comparing the results of the new model and the current method to the actual number of entrants from the CPS, we see in *Chart 7* that the newly specified model follows the CPS much more closely than the currently used method.

Chart 7. Entrants estimation methods compared to CPS entrants, United States, January 1996-December 2000



While the model produces robust estimates for this time period, it is reasonable to think that the current fixed parameters may not predict future estimates as accurately since the future predictions would be based on coefficients that were created using data from a fixed time period. As we move into the future, the relationships between the explanatory variables and the dependent variable may change over time. Evidence of the deterioration of these relationships can be seen in *Chart 7* as the model predictions begin to move further away from the CPS estimates toward the end of the forecasted period.

To correct this ongoing deterioration in the estimates, a procedure called a *rolling regression* was incorporated into the model. This procedure introduces new information into the model, as it becomes available, and disregards old information. The rolling regression adds a new month of data to the model while at the same time removing the oldest month's data. This technique enables the model to keep up with changing trends in the data, providing a more flexible approach to the estimation procedure.

The rolling regression technique was applied to forecast monthly entrants at the national level for January 1999 through December 2000. *Table 10* shows the entire monthly model parameters used to forecast this period. These parameters resulted from individual regressions that were run for each of the 24 forecasted periods using data observations for the previous 228 periods. For instance, to forecast January 1999, data observations from January 1980 through December 1998 were incorporated into the model and yielded the parameters listed in *Table 10* for January 1999. To predict February 1999, the oldest observations, which were for January 1980, were dropped and the latest observations for January 1999 were added and new parameters were developed. This procedure was repeated for each month up to December 2000.

Table 10. Monthly model parameters for rolling regression, January 1999-December 2000

Predicted Month	Jan-99	Feb-99	Mar-99	Apr-99	May-99	Jun-99	Jul-99	Aug-99	Sep-99	Oct-99	Nov-99	Dec-99
Intercept	0.002657	0.002723	0.002711	0.002577	0.00218	0.002426	0.002416	0.002343	0.001942	0.002066	0.002056	0.00152
IC	-0.4002	-0.4255	-0.4423	-0.4853	-0.4581	-0.5029	-0.4964	-0.4766	-0.4595	-0.4741	-0.4707	-0.5037
CC	0.0107	0.0127	0.0145	0.0199	0.0194	0.0204	0.0147	0.0195	0.0394	0.0373	0.0372	0.0526
Teens	0.0485	0.0469	0.0478	0.0493	0.0555	0.0505	0.0494	0.05	0.054	0.0526	0.0522	0.06
UR	0.001455	0.00147	0.001461	0.001465	0.001435	0.001476	0.001504	0.001489	0.001446	0.001455	0.001459	0.001425
D1	0.002407	0.002384	0.002389	0.0024	0.002394	0.00241	0.002419	0.002402	0.002362	0.002373	0.002374	0.002283
D2	0.001371	0.001359	0.001409	0.001448	0.001448	0.001467	0.001461	0.001424	0.001344	0.001369	0.001365	0.001296
D3	0.000744	0.000721	0.000738	0.000686	0.000729	0.000688	0.000692	0.000683	0.000634	0.000638	0.000641	0.000532
D4	0.000187	0.000167	0.000172	0.000135	0.000181	0.000136	0.000147	0.000141	0.0000992	0.000101	0.000107	-0.00000651
D5	0.002701	0.002695	0.002697	0.002665	0.002723	0.002613	0.002627	0.002625	0.002606	0.002607	0.002615	0.002518
D6	0.006072	0.006066	0.006065	0.006039	0.006037	0.005935	0.005876	0.005889	0.005916	0.005909	0.005918	0.005844
D7	0.003872	0.003864	0.003867	0.00385	0.003859	0.003809	0.003761	0.00376	0.003822	0.003811	0.003818	0.003772
D8	0.002419	0.002428	0.00244	0.002449	0.002438	0.002433	0.002406	0.002398	0.002382	0.002381	0.002387	0.002347
D9	0.002241	0.002235	0.002231	0.002214	0.002219	0.002195	0.002183	0.002181	0.002236	0.002214	0.002224	0.002174
D10	0.001945	0.001941	0.001937	0.001925	0.001931	0.001913	0.001902	0.00191	0.001922	0.001892	0.001879	0.001854
D11	0.001403	0.001401	0.001399	0.001392	0.001394	0.001386	0.001381	0.001389	0.001405	0.001399	0.001396	0.001364
Data Periods												
Start	**Jan-80**	**Feb-80**	**Mar-80**	**Apr-80**	**May-80**	**Jun-80**	**Jul-80**	**Aug-80**	**Sep-80**	**Oct-80**	**Nov-80**	**Dec-80**
End	**Dec-98**	**Jan-99**	**Feb-99**	**Mar-99**	**Apr-99**	**May-99**	**Jun-99**	**Jul-99**	**Aug-99**	**Sep-99**	**Oct-99**	**Nov-99**

Predicted Month	Jan-00	Feb-00	Mar-00	Apr-00	May-00	Jun-00	Jul-00	Aug-00	Sep-00	Oct-00	Nov-00	Dec-00
Intercept	0.001826	0.001484	0.001659	0.001512	0.000763	0.001404	0.001245	0.001108	0.000676	0.000708	0.001428	0.001759
IC	-0.5064	-0.5237	-0.5016	-0.4996	-0.537	-0.5161	-0.5307	-0.5346	-0.5162	-0.5366	-0.5348	-0.5089
CC	0.0502	0.0628	0.0584	0.0611	0.0793	0.0754	0.0704	0.0688	0.0682	0.062	0.0743	0.0716
Teens	0.0553	0.0593	0.0564	0.0584	0.0683	0.0595	0.0603	0.0617	0.0673	0.0651	0.0558	0.051
UR	0.001443	0.001423	0.001438	0.001427	0.001385	0.001403	0.001434	0.001441	0.001429	0.001472	0.001455	0.00146
D1	0.002274	0.002238	0.002233	0.002229	0.00222	0.002213	0.002231	0.002238	0.002234	0.002261	0.002233	0.002257
D2	0.001321	0.001309	0.001259	0.001256	0.001275	0.001238	0.00126	0.00127	0.001262	0.001292	0.00124	0.00124
D3	0.000554	0.000477	0.000458	0.000464	0.000445	0.000431	0.000437	0.000443	0.000461	0.000469	0.000419	0.000453
D4	0.0000174	-0.000057	-0.000047	-0.000037	-0.000092	-0.00011	-0.000099	-0.000091	-0.000071	-0.000057	-0.00011	-0.000067
D5	0.002546	0.002491	0.002501	0.002493	0.002497	0.002465	0.002483	0.002493	0.002508	0.002533	0.002497	0.002533
D6	0.005866	0.005833	0.005842	0.005839	0.005792	0.005736	0.00567	0.005679	0.005701	0.005712	0.005695	0.00573
D7	0.003785	0.00377	0.00377	0.003773	0.003763	0.003765	0.003718	0.003691	0.003721	0.003716	0.003707	0.003735
D8	0.002365	0.002347	0.002338	0.002335	0.00233	0.002324	0.002299	0.002292	0.002299	0.00232	0.002287	0.002299
D9	0.002189	0.002169	0.002177	0.002176	0.002155	0.002164	0.002141	0.002128	0.002165	0.002096	0.002064	0.002092
D10	0.001864	0.00186	0.001865	0.001867	0.001866	0.001873	0.001852	0.001843	0.001843	0.001798	0.001753	0.001775
D11	0.001371	0.001372	0.001375	0.001377	0.001379	0.001384	0.001374	0.001369	0.001369	0.00135	0.001295	0.001309
Data Periods												
Start	**Jan-81**	**Feb-81**	**Mar-81**	**Apr-81**	**May-81**	**Jun-81**	**Jul-81**	**Aug-81**	**Sep-81**	**Oct-81**	**Nov-81**	**Dec-81**
End	**Dec-99**	**Jan-00**	**Feb-00**	**Mar-00**	**Apr-00**	**May-00**	**Jun-00**	**Jul-00**	**Aug-00**	**Sep-00**	**Oct-00**	**Nov-00**

The forecasts for the out-of-sample observations for the time period of January 1999 through December 2000 that were produced using the entrants/population model with rolling regression were examined along side the forecasts for the same period produced from entrants/population model using fixed parameters and the current method.

The forecasts using the entrants/population model using fixed parameters for the period of January 1999 through December 2000 were generated using national data from January 1980 through December 1998. The model parameters are listed below in *Table 11*. Both versions of the entrants/population model were compared to the current method's estimates that were calculated for the same time period using equations 2 through 5 previously described in Section III of this paper.

Table 11. Entrants/Population Model Parameters for January 1999-December 2000 Forecasts, Generalized Least Squares

Variable	Parameter Estimate	Standard Error	t Value
Intercept	0.002226	0.001404	1.59
$(IC/Pop_{20+})_{t-1}$	-0.3724	0.2635	-1.41
$(CC/Pop_{20+})_{t-1}$	-0.0155	0.0674	-0.23
$(Pop_{16-19}/Pop_{20+})_{t-1}$	0.0519	0.019	2.73
UR_{t-1}	0.001519	0.000165	9.21
D_1	0.002399	0.000256	9.37
D_2	0.001367	0.000461	2.97
D_3	0.000727	0.000382	1.90
D_4	0.00025	0.000399	0.63
D_5	0.002706	0.000353	7.67
D_6	0.006205	0.000331	18.75
D_7	0.003828	0.00033	11.60
D_8	0.002474	0.000331	7.47
D_9	0.002301	0.000304	7.57
D_{10}	0.001972	0.00028	7.04
D_{11}	0.001416	0.00022	6.44

These three methods were then compared to the CPS entrants in *Chart 8*. Both models produced estimates that are much more inline with the CPS and far better than the current method. However, the estimates produced with the rolling regression follow the CPS slightly better. In time as more observations are added to the series, we can expect the model with fixed coefficient to become increasingly less accurate while the model incorporating the rolling regression technique will continue to produce quality estimates since the parameters are constantly updated with new monthly information.

Chart 8. Comparison of the entrant/population models to the current method and the CPS, United states, January 1999-December 2000

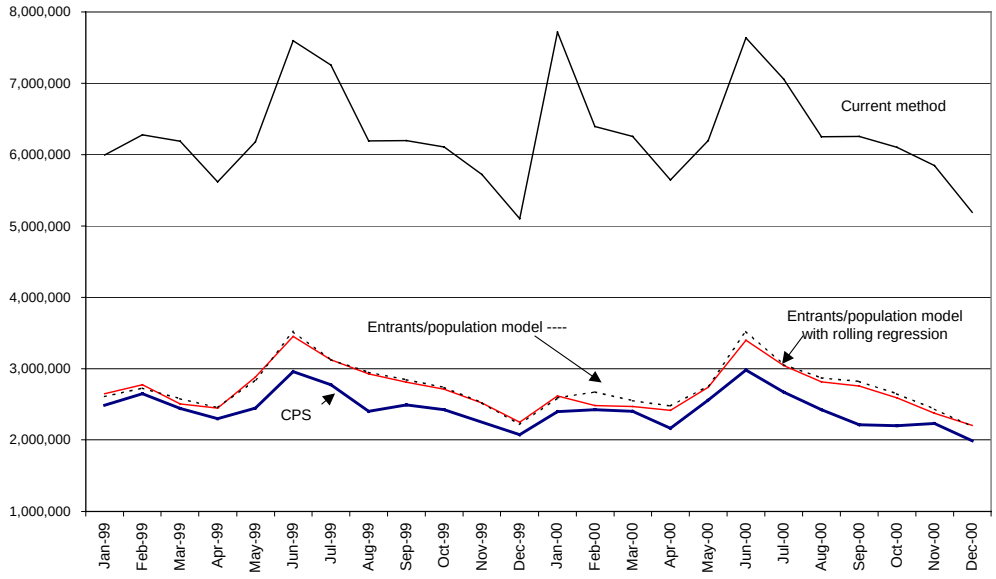


Table 12 shows the comparison of the forecasting errors, RMSE and MAE, for the entrants/population model estimates with and without the rolling regression to the current method. The entrants/population models have substantially lower forecast errors than the current method. The lower forecasting errors for the model incorporating the rolling regression indicate further improvement in the accuracy of the forecasts.

Table 12. Forecasting Errors, National Data, January 1999-December 2000

Methodology	RMSE	MAE
Current Method	3,892	3,859
Entrants/Population Model	335	304
Entrants/Population Model *with rolling regression*	311	276

Now that we have established that the better procedure for estimating unemployed entrants is the use of the entrants/population model with a rolling regression, we will use the model parameters developed in this section and apply them to state and labor market areas.

VIII. Application of the Model

The same procedures described in the previous section for testing the entrants/population model on the national level were applied to the state level. The model was tested by first forecasting the 60 out-of-sample periods of January 1996 through December 2000. Then the rolling regression technique was applied to 24 out-of-sample periods for January 1999 through December 2000.

Three states were used for testing: California, Florida, and Texas. These states were selected because they are among the four most populated states. The CPS sample sizes in these states are large enough to obtain statistically reliable monthly estimates for unemployed entrants into the labor force. CPS entrants data for these states were used to determine how well the current method and the new model predicted entrants.

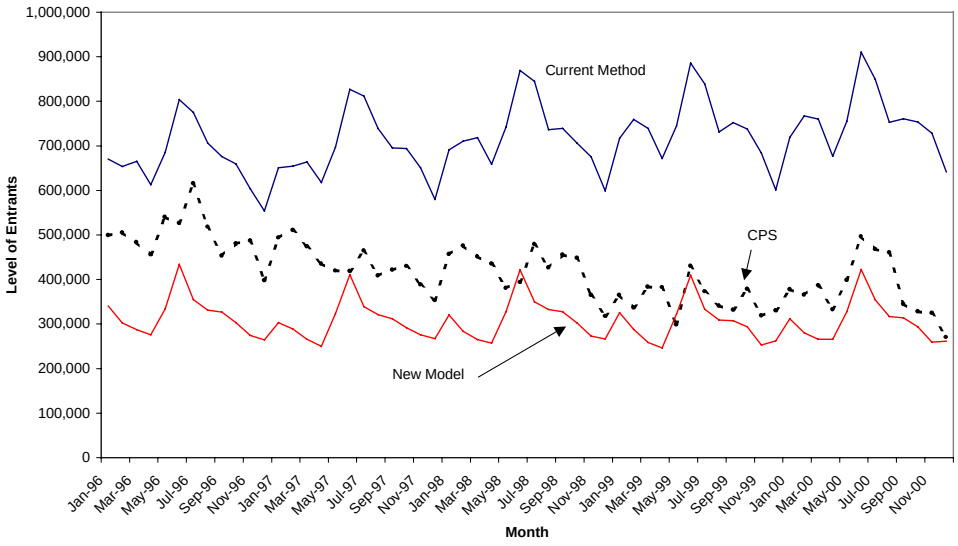
Individual regressions were not performed for each of these states. Instead the model parameters that were derived from using national data were applied to the state data. *(See Table 8.)* Like the original A and B factor models, the idea was to develop a single model using national observations that could be applied to the state and labor market levels. The same procedures that were used in comparing the entrants/population model to the current model for national data were used to test the model on the state level. Out-of-sample observations were used to predict monthly entrants for January 1996 to December 2000. The use of the new model at the state level proved successful. The state estimates produced by the model were an improvement over the estimates produced using the current method. The resulting forecast errors, the RMSE and MAE, for the entrants/population model were considerably smaller that they are for the current method *(See Table 13.)*.

Table 13. Forecasting Errors, State Data, January 1996-December 2000

Methodology	RMSE	MAE
California		
Current Method	313667	298139
Entrants/Population Model	128329	111783
Florida		
Current Method	175312	171228
Entrants/Population Model	28239	23129
Texas		
Current Method	240870	235313
Entrants/Population Model	39831	29794

Looking at the entrants estimates for California for period of January 1996 through December 2000 in *Chart 9*, we can see that monthly differences for the between the current method compared to the CPS was greater than the difference between the CPS and the model.

Chart 9 . Entrants estimation methods compared to CPS entrants, California, January 1996-December 2000



The model estimates produced similar results for Florida and Texas. *Chart 10* shows that the model predictions of entrants in Florida are more in line with the number of entrants in Florida from the CPS.

The comparison of estimates for Texas is displayed in *Chart 11*. Again it is clear that the model estimates are an improvement over the current method.

Chart 10. Entrants estimation methods compared to CPS entrants, Florida, January 1996-December 2000

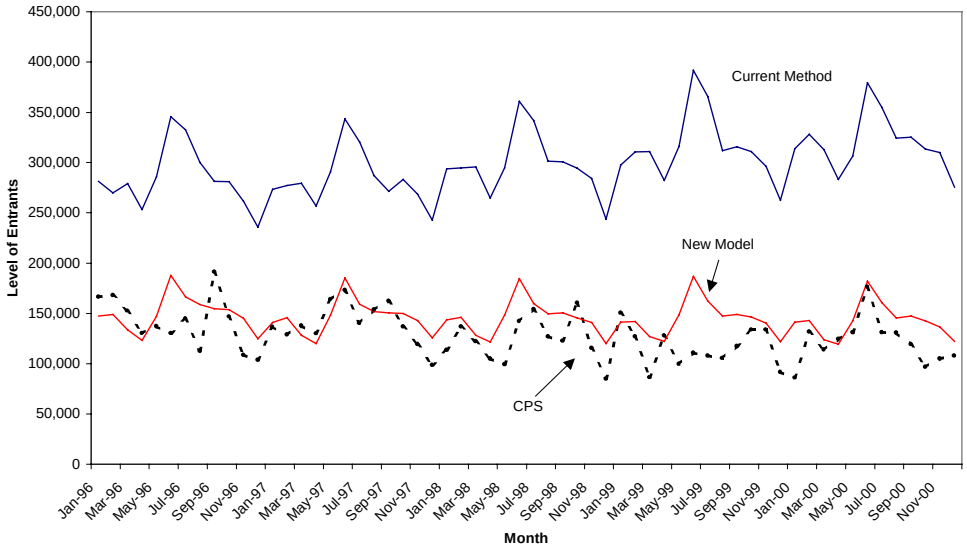
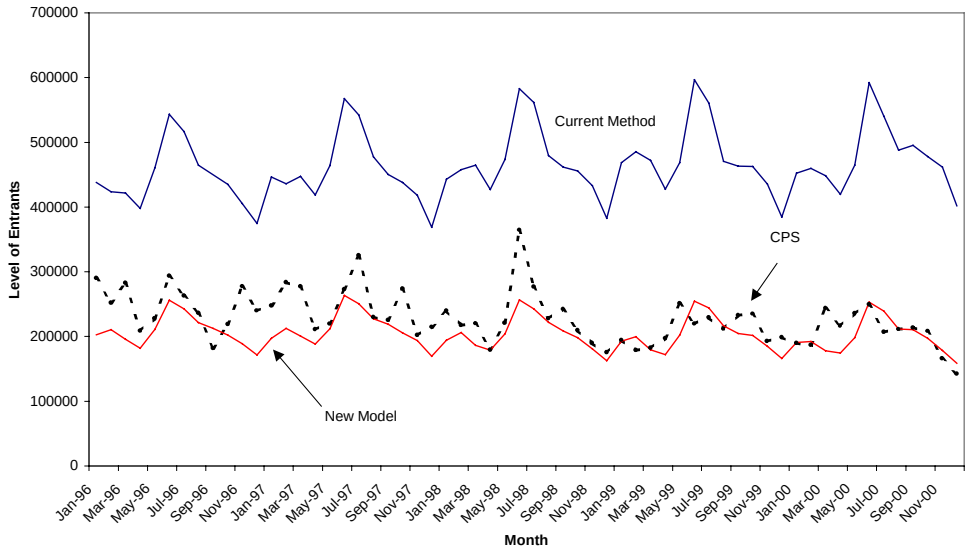


Chart 11. Entrants estimation methods compared to CPS entrants, Texas, January 1996-December 2000



Although the estimates produced by the entrants/population model are an improvement over the current method, further enhancements can be made to the estimates by application of the rolling regression technique.

The same rolling regression technique and parameters that were previously used for national data *(See Table 10. for model parameters.)* were applied to the state data for 24 out-of-sample observations for the period of January 1999 to December 2000. The forecasting errors are listed in *Table 14*. The results are similar to the errors measurements produced when using the model with the constant coefficients. (*See Table 13*.) The entrants/population model produced considerable lower values for the RMSE and MAE.

Table 14. Forecasting Errors, State Data, January 1999-December 2000

Methodology	RMSE	MAE
California		
Current Method	383068	379810
Entrants/Population Model *with rolling regression*	102511	93147
Florida		
Current Method	200600	198118
Entrants/Population Model *with rolling regression*	27389	21039
Texas		
Current Method	271038	266942
Entrants/Population Model *with rolling regression*	33430	26249

The use of the rolling regression to the state data improved the predictions of entrants and brought them more in line with the actual CPS entrants. A comparison of the CPS entrants to the current method and the entrants/population model with rolling regression is depicted for the test states in *Charts 12-14*.

Chart 12. Comparison of entrants/popoulation model with rolling regression to the current method and the CPS, California, January 1999-December 2000

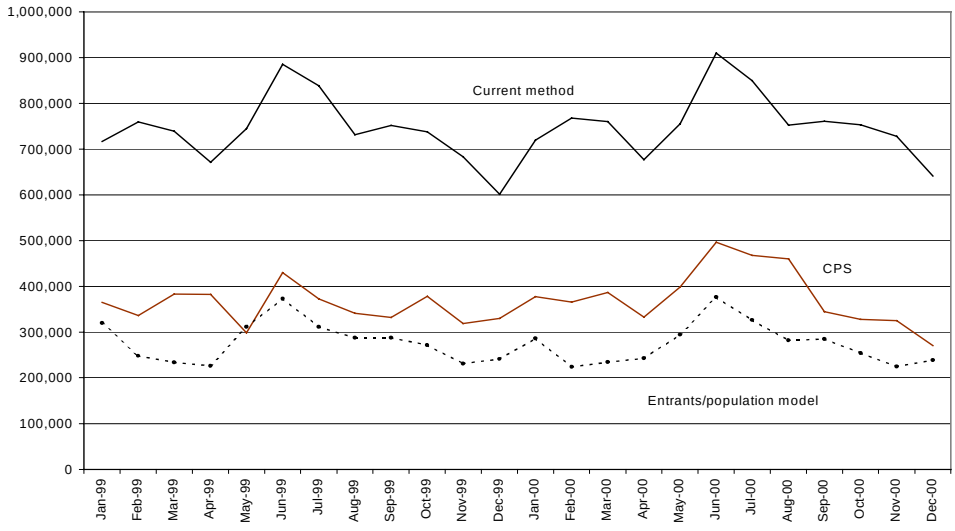


Chart 13. Comparison of entrants/popoulation model with rolling regression to the current method and the CPS, Florida, January 1999-December 2000

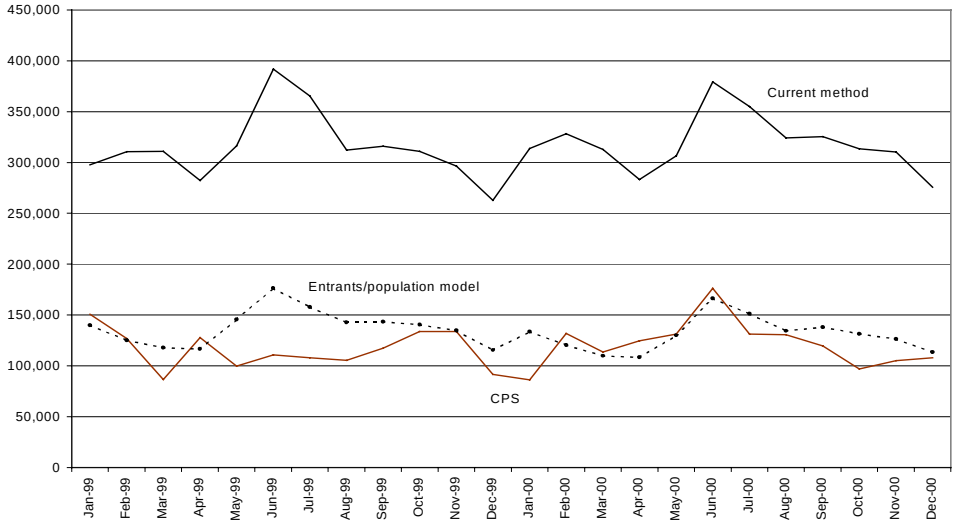
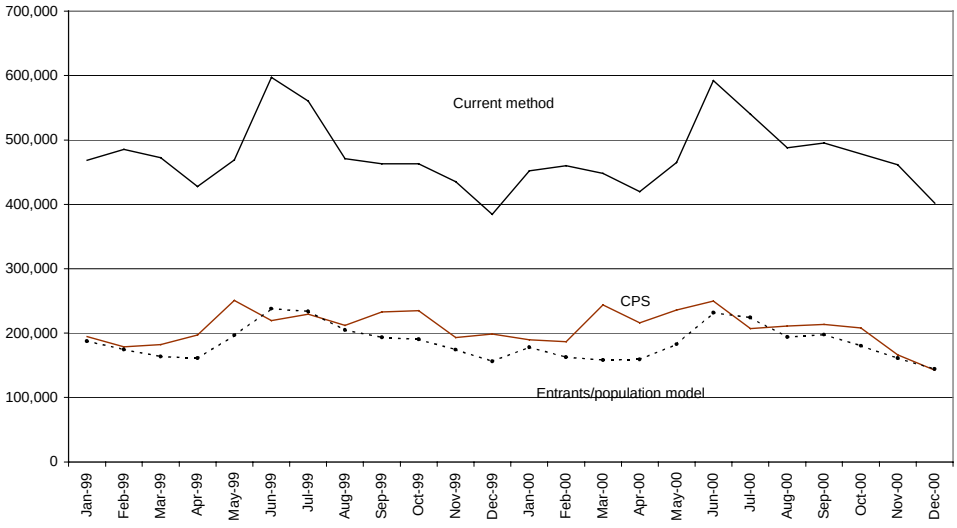


Chart 14. Comparison of entrants/popoulation model with rolling regression to the current method and the CPS, Texas, January 1999-December 2000



The next step was to test the model on sub-state estimates. Since no monthly CPS entrants data exist for labor market areas, the only measure of improvement is a comparison of the new handbook estimates produced using the entrants/population model to the actual handbook estimates, described *in Section II*, and to the official statewide estimates.

In *Section II*, information was provided to explain how the handbook method calculates unemployment estimates at the labor market area level. Recall that the unemployment estimates consist of continued claims, final payments, unemployed entrants into the labor force. The continued claims and final payments data are obtained from a sound and reliable source, which is the state UI system. It is the estimation of entrants that is the weak point in the handbook procedure.

The entrants/population model was applied to New Jersey monthly unemployment data for 1999 using the rolling regression model parameters in *Table 10* that were developed with national data. The model output for each month was entered into the handbook calculations replacing the individual A and B factor equations used in the current method of estimating entrants (*See pages 9 and 10*). Handbook estimates were produced in this manner for all labor market areas in New Jersey. The handbook unemployment estimates calculated using the entrants/population model yielded better estimates than the current method. *Table 15* lists these results along with the actual handbook unemployment estimates and the official statewide estimates.

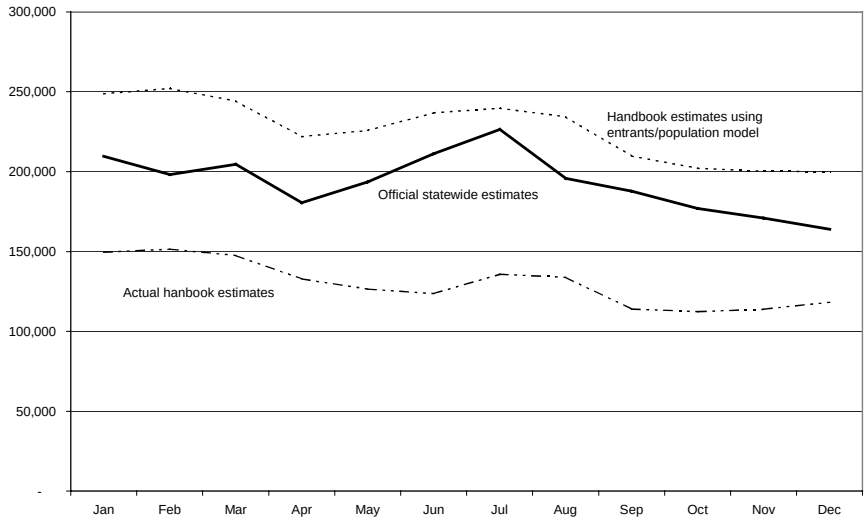
Table 15. Comparison of New Jersey unemployment estimates, January-December 1999

Month	Statewide Estimates	Actual Handbook Totals (Current Method)	Handbook Totals using Entrants/Population Model
Jan	209,713	149,611	248,790
Feb	198,227	151,492	252,151
Mar	204,491	147,535	244,204
Apr	180,541	132,881	221,958
May	193,538	126,526	225,849
Jun	211,248	123,680	236,849
Jul	226,401	135,801	239,704
Aug	195,740	133,805	234,371
Sep	187,821	114,017	209,813
Oct	177,037	112,339	202,092
Nov	170,873	113,936	200,245
Dec	163,970	118,322	200,157

Chart 15 graphs the total monthly unemployment levels resulting from the actual handbook method, the handbook method using the entrants/population model, and the official statewide unemployment estimates for New Jersey in 1999. The dashed line represents the official statewide unemployment estimates. The solid line represents the sum of the labor market areas unemployment estimates produced using the entrants/population model. The longer dashed line represents the sum of the labor market areas using the current method.

The unemployment estimates, produced with the entrants/population model, are not only closer to the statewide estimates, but they follow the trend of the statewide estimates more accurately. The average difference of the monthly estimates produced using the entrants/population model was approximately 33,000 greater than the official statewide estimates. The current method produces an average difference that was 63,000 lower than the official estimates.

Chart 15. Comparison of handbook LMA sums to official statewide unemployment estimates, New Jersey, January-December 1999



The unemployment additivity ratios at the state level were also used as a measure of how well the two entrants procedures estimated handbook unemployment compared to the official estimates. The additivity ratio is the proportion of the sum of the labor market areas' handbook estimates to the official statewide estimate. A ratio of 1.0 would indicate the two estimates match perfectly. If the ratio is greater than 1.0, this indicates that the handbook procedure is under estimating unemployment and a ratio less than 1.0 indicates overestimation.

The monthly sums of the individual labor market estimates from both procedures are divided by the official statewide estimates and listed in *Table 16*. The unemployment additivity ratios were reduced considerably.

Table 16. Comparison of New Jersey unemployment ratios, January-December 1999

Month	Actual Ratio	Ratio using entrants/population model
Jan	1.401722	0.842933
Feb	1.308498	0.786143
Mar	1.386051	0.837378
Apr	1.358667	0.813401
May	1.529630	0.856934
Jun	1.708021	0.891911
Jul	1.667153	0.944501
Aug	1.462875	0.835171
Sep	1.647307	0.895182
Oct	1.575918	0.876023
Nov	1.499728	0.853320
Dec	1.385795	0.819205

Although incorporating the entrants/population model into the Handbook estimates slightly overstates the Handbook unemployment; these estimates are still an improvement over the estimates produced using the current method. The estimates are closer to the official statewide estimates and follow trend more accurately.

Further testing of the entrant/population model can be undertaken by applying the additivity and disaggregation processes (see pages 7-8) to the handbook estimates produced with the model and compare these estimates to the official LAUS estimates. It is likely that unemployment will be redistributed within a state based on new information incorporated into the model. If this is the case, the local economic situations must be investigated to determine the reasonability of the new estimates.

IX. Implications of Results

Use of the entrants/population model developed in this paper has shown to be more accurate in estimating the number of unemployed entrants in to the labor force than the methodology that is currently in use. The model has been tested on three geographic levels for which the Bureau of Labor Statistics produces estimates; the nation, the states, and labor market areas.

The entrants component is a key part of the unemployment estimation procedure for labor market areas. It is also the weakest part of the procedure. The current method has proved to be an ineffective tool at estimating this component. Use of the entrants/population model in the production of labor force estimates produces more accurate measurements on the labor market area level. Improved quality of the estimates provides better information on the economic conditions of the labor market area.

The importance of producing quality labor force estimates is greater when one considers the number of Federal programs that distribute grants and entitlements based on these estimates, especially at the sub-state level. In fiscal year 2001, over 20 Federal programs used these estimates to allocate approximately $25 billion worth of funding to state and local governments. If labor force estimates are not accurate, areas in need of Federal assistance may not be getting the amount of funding they are need to remain economically stable in times of down turns and additional funding may go to areas where it is not needed. This can negatively impact major social programs that are designed to assist the needy based on the unemployment rate or level of unemployment.

Reference:

U.S. Department of Labor, Bureau of Employment Security. *Handbook on Estimating Unemployment*, Employment Security Research Methods, Handbook Series (BLS Reprint No. R-185), 1960.

U.S. Department of Labor, Bureau of Labor Statistics, *How the Government Measures Unemployment*, Report 864, 1994.

U.S. Department of Labor, Bureau of Labor Statistics, *Manual for Developing Local Area Unemployment Statistics*, December 1982.

U.S. Department of Labor, Bureau of Labor Statistics, *Local Area Unemployment Statistics Program Manual*, August 2001.

U.S. Department of Labor, Bureau of Labor Statistics, *BLS Handbook of Methods*, Bulletin 2490, April 1997.

U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, Economic and Statistics Administration, U.S. Census Bureau, *Current Population Survey Design and Methodology*, Technical Paper 63, March 2000.

Glossary

Additivity:	Procedure used to force the independent estimates for labor market areas, also known as handbook estimates, to add to the official statewide estimates.
A factor:	The factor that imparts the influence of the experienced labor force on that portion of the handbook estimate covering total entrant unemployment.
All other employment:	This category includes the self employed, unpaid family workers and domestics working in private households.
B factor:	The factor that imparts the influence of the experienced unemployed on that portion of the handbook estimate covering total entrant unemployment.
Base period:	A specific period of 12 consecutive months, or in some states, 52 weeks preceding the beginning of a benefit year during which an individual must have the required employment and/or wages in order to establish entitlement to compensations or allowances under a UI program.
Benefit year:	A period, generally 52 weeks, during which an individual claimant may receive their maximum potential benefit amount.
BLS:	Bureau of Labor Statistics.
CES:	Current Employment Statistics.
Civilian labor force:	The pool of available workers. It includes all civilians, 16 years and older who are either employed or unemployed.
Continued claims:	This type of claim represents a continuous spell of unemployment for individuals who meet UI eligibility requirements.

Covered employment:	Those jobs covered by the unemployment compensation programs. Does not include some agricultural workers, employees of religious and nonprofit organizations, household workers, and self-employed workers.
CPS:	The Current Population Survey is a monthly survey of approximately 60,000 households.
Employed:	Persons, who during the CPS reference week of the 12th of the month, did any work as paid employees, in their own business, or who worked at least 15 hours as unpaid workers in a family business are employed. Also included are persons who had a job or business, but were temporarily absent because of illness, vacation, bad weather, labor management disputes or personal reasons.
ETA:	The Employment and Training Administration is responsible for overseeing the UI system.
Entrants:	The sum of new entrants and reentrants into the labor force from the monthly Current Population Survey.
Experienced unemployed:	Unemployed persons receiving UI benefits or continued claims count. Defined as unemployed minus entrants.
Experienced labor force:	The sum of employed and the experienced unemployed.
Final payment:	The last payment to a claimant, which exhausts the individual's maximum potential benefit entitlement in the UI program. Also referred to as an *exhaustee*.
Handbook method:	The methodology used to create labor market area estimates. The handbook uses data from the CES survey, UI claims, and other sources, plus relationships between these data developed over time, to create the labor force estimates for labor market areas.

Handbook share	The ratio of the area's handbook estimate to the sum of the handbook estimates for all labor market areas in the state. This ratio is then applied to the official statewide estimate to produce the final estimates for each labor market area.
Initial claims:	Any notice of unemployment file to request a determination of entitlement to and eligibility for compensation.
Labor force:	The sum of employed and unemployed. Represents person with jobs and those who want jobs.
Labor Market Area:	Labor Market Area (LMA) is an economically integrated unit within which workers may readily change jobs without changing their place of residence. All states are divided into exhaustive labor market areas that usually include a county or a group of contiguous counties. (Except in New England, where cities and towns are used in place of counties.) Independent Handbook estimates of employment and unemployment are made monthly for each labor market area and form the basis for the LAUS estimates.
LAUS:	Local Area Unemployment Statistics.
LAUS estimate:	The BLS published labor force estimates derived in one of three ways: by a regression model approach for the state level; based on Handbook estimates adjusted to the state level; or disaggregated from a labor market area.
New entrants:	In the CPS, new entrants are prospective workers looking for a job, such as students entering the labor force after graduation from school, and others who have not previously worked.
Unemployment Rate:	Major economic indicator from the CPS. Defined as unemployed divided by the labor force multiplied by 100.

Payroll employment:	Officially named Current Employment Statistics, it is a monthly survey conducted by BLS that counts the number of jobs.
Reentrant:	Unemployed individuals who have previously worked but were out of the labor force prior to their most recent job search.
Unemployed:	Persons who had no employment during the CPS reference week of the 12th but were available for work, except for temporary illness, and made specific efforts to find work during the preceding 4 weeks are unemployed.
Unemployment rate:	The number of unemployed persons expressed as a percent of the civilian labor force.
UI:	Unemployment insurance.
YPR:	The youth population ratio is the population of 16-19 age group divided by the population of the 20 and over age group.

Walter J. Sylva Jr.

Academic experience:

Masters of Art, Economics,
Virginia Polytechnic Institute and State University
Falls Church, Virginia
2002

Bachelor's of Art, Economics,
Montclair State University
Upper Montclair, New Jersey
1989

Work experience:

Economist
United States Department of Labor,
Bureau of Labor Statistics
Washington, DC
1990-1998, 2000-present

Industry Economist
United States Department of the Army,
Military Traffic Management Command
Falls Church, Virginia
1999-2000