V. PRELIMINARY EMPIRICAL ANALYSIS BASED ON AVAILABLE DATA

V. (a) EXISTING RHODE ISLAND NURSING DATA

At the time work for this grant commenced, the existing time series data on nurses in Rhode Island, was, to put it kindly, sparse. This is not to suggest that Rhode Island was the only state with such a paucity of existing data. It did, however, severely curtail the types of empirical analysis that could be performed "out of the gate." Unfortunately, it will likely take several years to amass the type of time series dataset for the entire state that is needed to conduct policy simulations that will be of greatest utility when making decisions about actions such as changing the wages of CNA's in this state.

In spite of such data limitations, we were able to move forward and get as much out of the existing time series data as is possible. The first step consisted of "filling in the gaps" with existing data. Equipped with this new dataset, some empirical analysis was performed, most notably trend analysis. The following sections detail the data that existed at the beginning of this grant, the data that was added, the results of trend analysis and its implications, and existing information that was added at the state level, pertaining to licensed CNA's.

EXISTING EMPLOYMENT AND WAGE DATA

Data on nurse employment and wages did exist at the time work on this grant began, but for the late 1980s and early 1990s, surveys were only taken every three years. And, for 1989, 1992, and 1995, the published employment data pertained only to hospitals. Consecutive annual data on total nursing employment began in 1996 and extended through 2000. As the table below shows, for Nurse Aides Orderlies and Attendants (NAOA's), employment data for the year 1999 was not published.[1] As should be clear

EXISTING EMPLOYMENT DATA			
	RN	**LPN**	**NAOA**
1996	**9,908**	**2,580**	**8,045**
1997	**9,130**	**2,260**	**8,530**
1998	**9,770**	**2,410**	**8,710**
1999	**11,290**	**2,440**	**----**
2000	**11,130**	**2,080**	**7,560**

from this data, only five observations were available for RN's and LPN's, and four observations existed for NAOA's. The data is still useful, in that it shows some potentially important trends through time. LPN employment, for example, has been trending down over this period, a behavior similar to that for LPN's nationally. In contrast, RN employment has been rising, in general, from 1996 through 1999, while 2000 data indicates a slight decline. The NAOA data, however, are potentially troubling, as a pronounced upward trend exists from 1996 through 1998, then, after a missing value, 2000 NAOA employment is significantly below the level for any of the years preceding it.

[1] Apparently, the data for this year failed to meet the reliability requirements of the US Bureau of Labor Statistics. They did, however, publish wage data for NAOA's that year.

The existing data for nurse wages was similar to, but in two ways, better than the employment data. The similarity is that consistent data existed only for the time period from 1996 through 2000. The two ways in which this data appears to be better than that for employment are that there are no missing values for NAOA's,[2] and the wage trends displayed by this data appear to be consistent for all three nurse categories, unlike employment, where a dramatic reversal in employment appears after the missing observation (in the year 2000).

EXISTING WAGE DATA			
	RN	**LPN**	**NAOA**
1996	**$19.52**	**$15.65**	**$ 7.89**
1997	**$20.81**	**$16.02**	**$ 8.68**
1998	**$22.63**	**$16.09**	**$ 8.94**
1999	**$23.10**	**$16.77**	**$ 9.31**
2000	**$22.88**	**$18.59**	**$10.16**

Two of the data points stand out in this table. First, for RN's, the average hourly wage dropped in 2000, accompanying a slight decline in employment. Second, a dramatic rise in LPN wages (+10.9 percent) occurs in that year, coinciding with a substantial decline in LPN employment (by almost 15 percent). This is an interesting contrast in behavior, especially since 2000 was the last fully year of the previous recovery, where labor markets were very tight and Rhode Island was at full employment.[3] Unfortunately, nothing can be inferred about the possibility that NAOA's were substituted for either RN's or LPN's in 2000 since the missing value in 1999 NAOA employment precludes us from knowing whether employment for that nurse category was actually higher or lower than its value in 1999.

FILLING IN THE GAPS

Even though actual statewide data does not exist for the overall nurse employment (hospital and other sources) prior to 1996 and consecutive values of historical wage data are not available prior to 1996, it is possible to use econometric methods to simulate these values.

In order to accomplish this, actual (known) values of nurse employment or wages for each nurse category are used as the dependent variable in a multiple regression equation with explanatory variable(s) whose values extend back to the desired period that will be simulated. Predicted values from these equations provide the values for the missing time periods, thus allowing a consistent set of historical data to be obtained.[4] Employment

[2] The fact that wage data for NAOA's exists for 1999 while employment data fail to satisfy BLS quality requirements is rather curious, since employment and wages are functionally dependent on each other.

[3] Rhode Island's unemployment rate in 2000 was 4.1 percent (Rhode Island Department of Employment and Training), which is below the threshold of 4.2 percent where full employment exists.

[4] Actual values of either employment or wages are used for the years 1996 through 2000.

values were generated avoiding wage variables since this data will ultimately be used to econometrically link employment to wages. Had wages been used in those equations, the simulated employment data would have had a built-in correlation with wages. Similarly, wages were not simulated using employment data.

The time period for which this simulation was performed extends back to 1988. While technically, it is possible to generate values for years prior to 1988, there is a problem in that the underlying structure of Rhode Island's economy changed in late 1987, moving from a manufacturing-based economy to a service and information-based economy.[5] This potentially introduces a number of complications for the statistical analysis. In light of this, it was decided that the most accurate and timely estimates of employment and wages would be derived for the time period from 1988 through the present.

The variables included in each of these equations are summarized below:[6]

RN's

Employment = f (national RN employment, dichotomous variable for 1997)
Wages = f (current wage, lagged wage, dichotomous variable for 1998)

LPN's

Employment = f (national LPN employment, simulated RN employment in RI as a percentage of RI payroll employment, dichotomous variable for 2000)
Wages = f (weekly earnings of LPN's nationally, dichotomous variable for 1999)

NAOA's

Employment = f (RI unemployment rate, health service employment in RI as a percentage of RI payroll employment)
Wage = f (average manufacturing wage in RI, dichotomous variable for 2000)

---

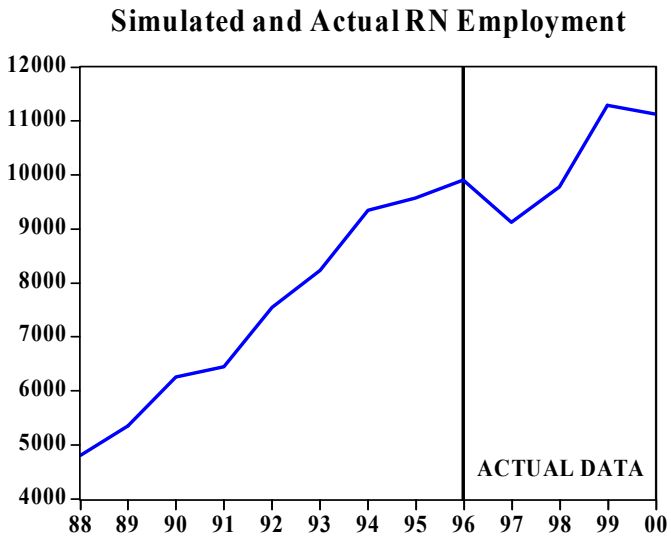
[5] This is documented in detail in Dr. Lardaro's Internet web site, "Transition to a Service Economy," **http://members.cox.net/lardaro/transition.htm .**

[6] For notational purposes only, each equation uses f( ) to denote that the dependent variable is related to the variables contained in the parenthesis. *The same notation of f( )for each equation should not be taken to indicate that the exact same function was used for each.*

RESULTS

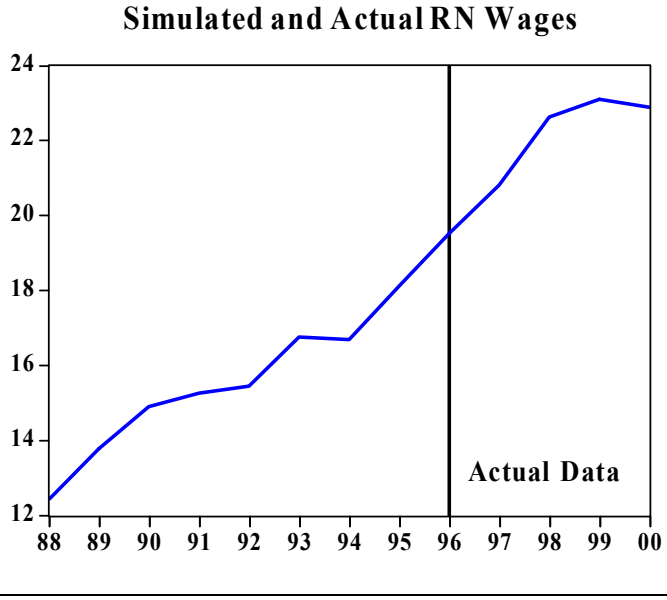
RN EMPLOYMENT

RN Employment	
1988	4,800.1
1989	5,338.7
1990	6,262.1
1991	6,449.0
1992	7,548.3
1993	8,229.8
1994	9,340.0
1995	9,570.9
1996	9,908.0
1997	9,130.0
1998	9,770.0
1999	11,290.0
2000	11,130.0



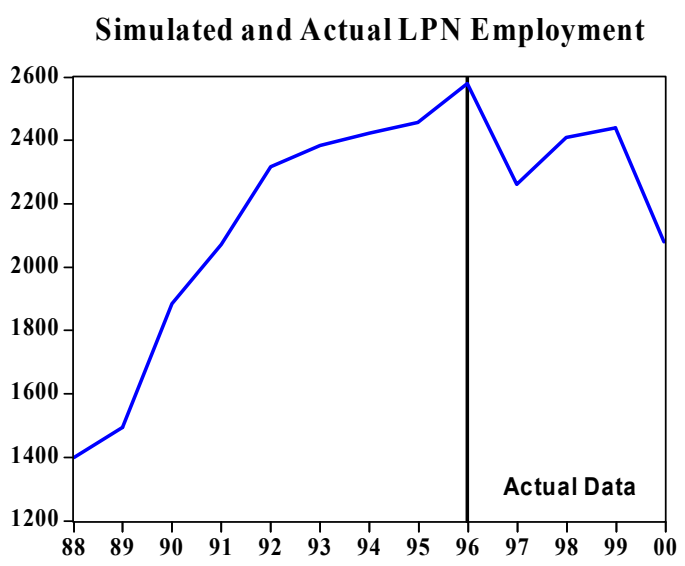
RN WAGES

RN Wages	
1988	$12.45
1989	$13.77
1990	$14.90
1991	$15.26
1992	$15.45
1993	$16.77
1994	$16.69
1995	$18.13
1996	$19.52
1997	$20.81
1998	$22.63
1999	$23.10
2000	$22.88



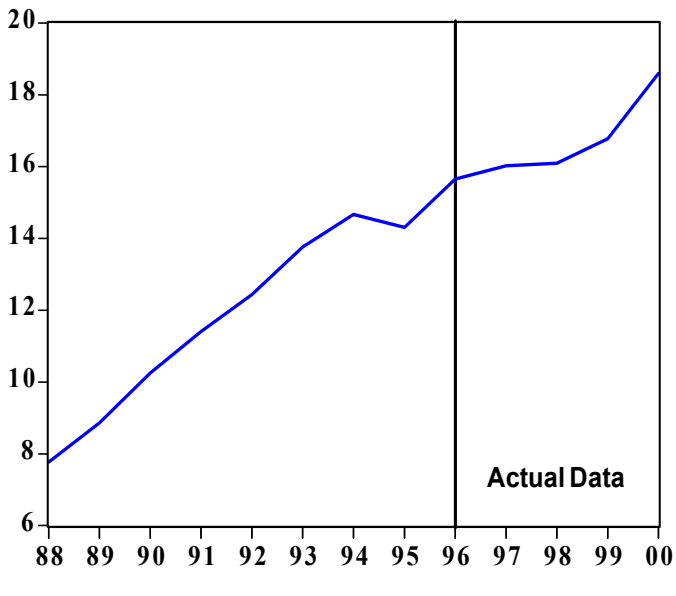
LPN EMPLOYMENT

LPN Employment	
1988	1,398.1
1989	1,493.2
1990	1,884.3
1991	2,072.4
1992	2,316.1
1993	2,384.2
1994	2,422.8
1995	2,455.6
1996	2,580.0
1997	2,260.0
1998	2,410.0
1999	2,440.0
2000	2,080.0



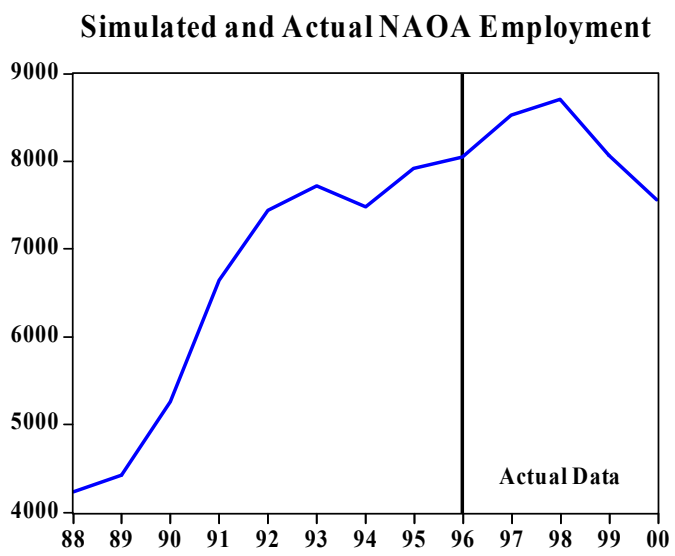
LPN WAGES

LPN Wages	
1988	$7.77
1989	$8.86
1990	$10.25
1991	$11.40
1992	$12.43
1993	$13.76
1994	$14.67
1995	$14.30
1996	$15.65
1997	$16.02
1998	$16.09
1999	$16.77
2000	$18.59



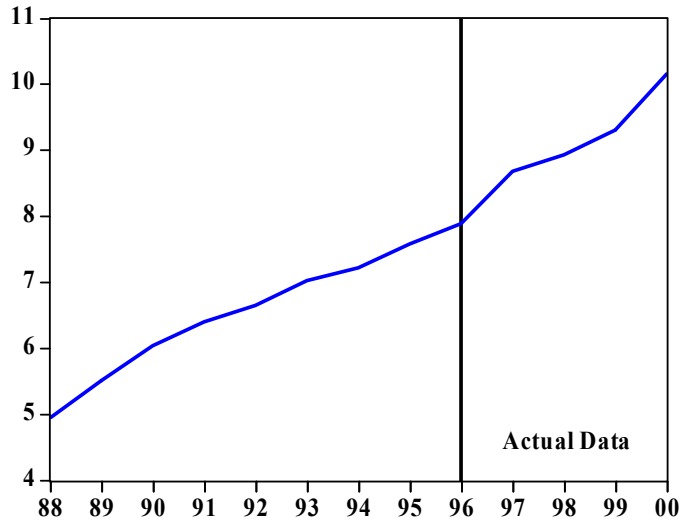
NAOA EMPLOYMENT

NAOA Employment	
1988	4,236.8
1989	4,421.1
1990	5,256.8
1991	6,649.1
1992	7,440.4
1993	7,715.6
1994	7,481.2
1995	7,924.3
1996	8,045.0
1997	8,530.0
1998	8,710.0
1999	8,068.5
2000	7,560.0



NAOA WAGE

NAOA Wages	
1988	$4.95
1989	$5.52
1990	$6.04
1991	$6.41
1992	$6.66
1993	$7.02
1994	$7.23
1995	$7.58
1996	$7.89
1997	$8.68
1998	$8.94
1999	$9.31
2000	$10.16



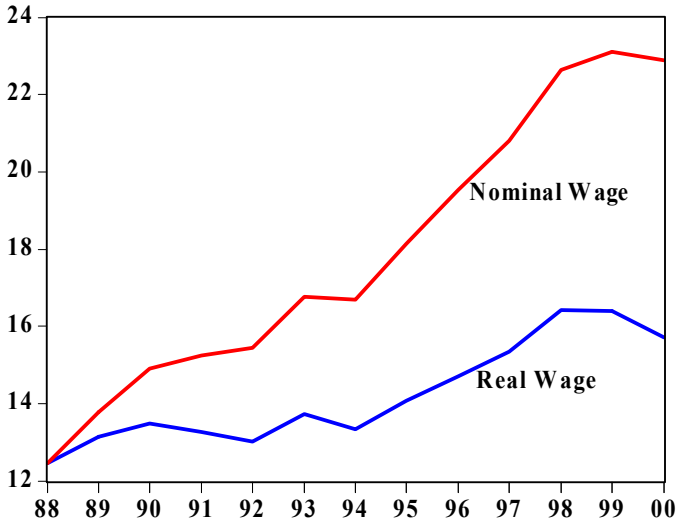
NOMINAL AND REAL WAGES

Equipped with time series data on hourly wages extending back to 1988, it is meaningful to calculate the inflation adjusted, or real wages of nurses in Rhode Island. This is accomplished using the Consumer Price Index for All Urban Consumers (CPI_U). In order to make the resulting numbers more comprehensible to the intended audience of this report, the base year was changed to the first year of the Rhode Island nursing data, 1988. In this way, nominal and real wages are identical in the first year of the data, which gives a common starting point.

The real wage is critical for both labor demand and labor supply decisions, since it is a relative magnitude, expressed as nominal wage relative to prices. Higher or rising prices tend to be consistent with greater labor demand since, given the existing level of wages, product prices rise relative to what firms are paying workers, causing labor to become relatively less expensive. The opposite is true for workers. Higher prices erode the purchasing power of nominal wages, lowering the real wage workers receive. Rising prices tend to discourage labor supply (other things being equal), since the real return to hours worked falls, making work hours relatively less attractive than time spent in non-work activities.

Registered Nurses in Rhode Island have seen their nominal wages rise by 83.8 percent over the period from 1988 through 2000. The estimated average annual growth rate in annual RN wages is 5.3 percent.[7] As the table shows, however, while real wages have also grown through time, their total rise was only 31.3 percent nominal wage growth.

	RN Wages	
	Nominal	Real
1988	$12.45	$12.45
1989	$13.77	$13.14
1990	$14.90	$13.48
1991	$15.26	$13.25
1992	$15.45	$13.03
1993	$16.77	$13.73
1994	$16.69	$13.32
1995	$18.13	$14.07
1996	$19.52	$14.72
1997	$20.81	$15.34
1998	$22.63	$16.42
1999	$23.10	$16.40
2000	$22.88	$15.72
% chg	83.8%	26.3%



[7] This value was estimated econometrically based on a constant growth rate model: $w_t = w_0(1+g)^t$, where **w** is the relevant wage rate, w_o is the initial wage, and **g** is the constant annual growth rate. The coefficient of the time trend was statistically significant in this and all of the growth rate calculations that follow.

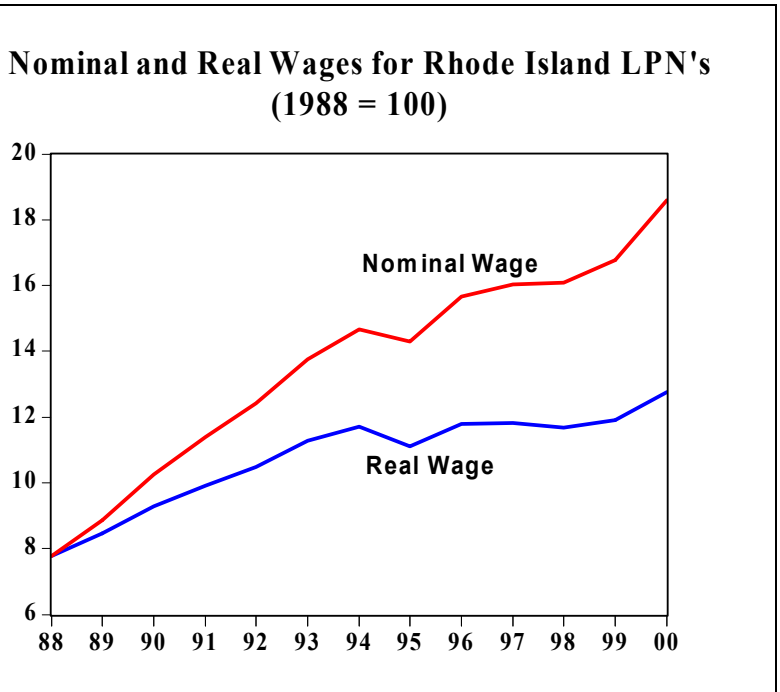
In fact, the estimated annual rate of growth in real wages for RN's is 2.2 percent, well below half the annual growth in nominal wages.

It is tempting to immediately relate this finding to the shortage of RN's in this state. The relatively slow growth in the real wages of RN's in this state may well be a fundamental factor underlying these shortages. But, based on the available data up to this point, nothing concrete can be concluded, for two reasons. First, the time series employment data that is now available is neither strictly demand data nor supply data (see below), but an unknown combination of both. As such, it is not possible to state with any degree of confidence that labor supply has been adversely affected by the observed relatively slow growth in real wages, or that labor demand has been dramatically higher with slowly growing real wages. Second, before we can meaningfully establish an empirical link between the real wage and employment through time, other factors affecting *both* labor demand and labor supply must be taken into account. The information here is that of a simple (zero order) correlation. It is not possible to statistically establish causation.

So, we now have a potentially important tool with which to explain ongoing nurse labor shortages in this state. While real wages may very well prove to be a statistically significant explanatory factor of nursing shortages, it must be recognized that we have progressed only up to the first step in making this determination.

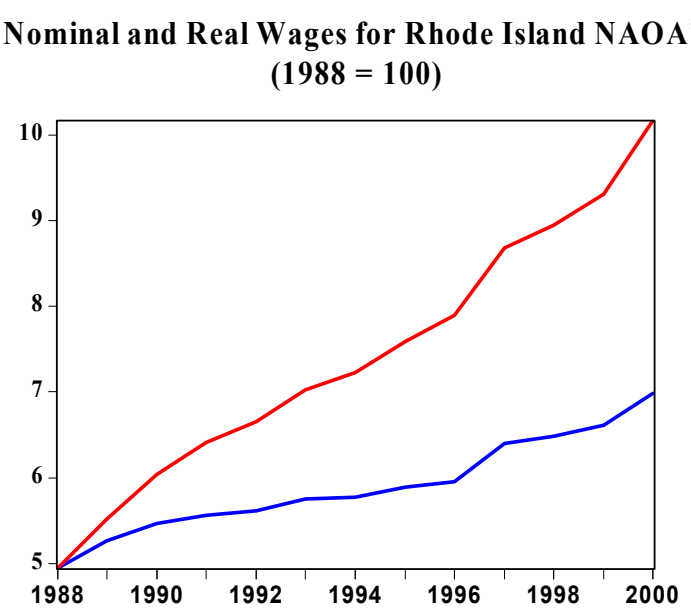
The situation for LPN's in Rhode Island is, in many ways, similar to that for RN's. Nominal wages have grown significantly over the 1988 – 2000 period, by 139.2 percent, but real wage growth approximately half that much (+ 64.3 percent). However, estimated average growth rates in both types of wages are higher for LPN's: 6.7 percent for nominal wages and 3.6 percent for real wages.

	LPN Wages	
	Nominal	Real
1988	$7.77	$7.77
1989	$8.86	$8.45
1990	$10.25	$9.28
1991	$11.40	$9.90
1992	$12.43	$10.48
1993	$13.76	$11.26
1994	$14.67	$11.71
1995	$14.30	$11.10
1996	$15.65	$11.80
1997	$16.02	$11.81
1998	$16.09	$11.68
1999	$16.77	$11.91
2000	$18.59	$12.77
% chg	139.2%	64.3%

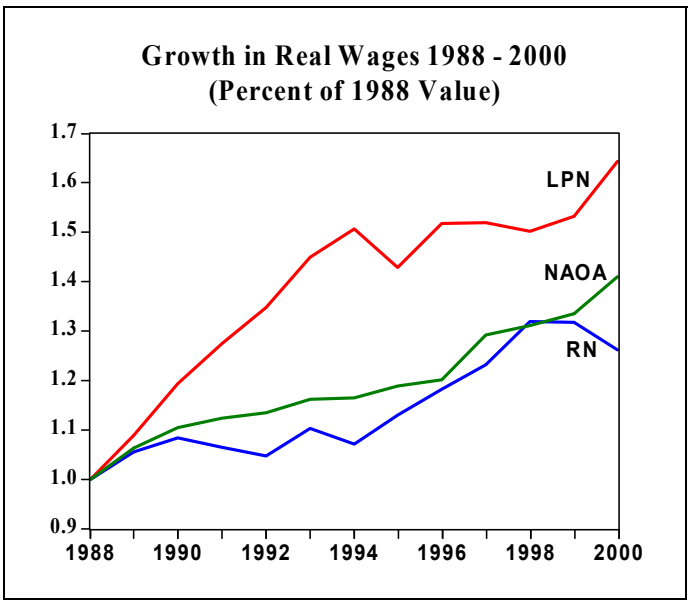


Finally, the historical behavior of nominal and real wages for NAOA's is between that of RN's and LPN's. Over the 1988 – 2000 period, nominal wages roughly doubled, rising by 105.3 percent. Real wages increased by 41 percent over this period, which is 39 percent of the gain in nominal wages. The estimated annual growth rate in nominal wages for NAOA's is 5.6 percent. The corresponding estimate for real wage growth is 2.5 percent.

	NAOA Wages	
	Nominal	Real
1988	$4.95	$4.95
1989	$5.52	$5.26
1990	$6.04	$5.46
1991	$6.41	$5.56
1992	$6.66	$5.61
1993	$7.02	$5.75
1994	$7.23	$5.77
1995	$7.58	$5.89
1996	$7.89	$5.95
1997	$8.68	$6.40
1998	$8.94	$6.49
1999	$9.31	$6.61
2000	$10.16	$6.98
% chg	105.3%	41.0%



The graph below provides a more direct comparison of real wages for each nurse class over the 1988 – 2000 period, expressing actual real wages for each year relative to their value in 1988. Not only has the growth in real wages for RN's consistently lagged behind



that of the other two nurse categories, but their level of real wages was the only one to decline in the year 2000. The table below summarizes the wage data.

Percent Change 1988 – 2000		
	Nominal	**Real**
RN	**83.8**	**26.3**
LPN	**139.3**	**64.4**
NAOA	**105.3**	**40.8**

- Real Wage growth substantially lags nominal wage growth for each nurse category
- The greatest real wage growth occurred for LPN's, the group whose employment has been declining for years
- For RNs, real wage growth is less than one-third of nominal wage growth and the lowest of any of the three nursing categories

TREND ANALYSIS

The next step in analyzing the complete time series dataset is trend analysis. Trend analysis is a time series technique sometimes used to forecast future values of a variable of interest. Several variations of this exist. Linear trend analysis, which many persons are familiar with, estimates a constant rate of change in a variable through time. Forecasts using this technique presuppose that this rate of change will remain valid in all future time periods. Other variants of trend analysis exist. Quadratic trend analysis models the time trend itself as a function of time: the rate of change in the variable of interest thus differs depending on the time period considered. This technique is appropriate when the data has an inflection point, where the series becomes steeper or flatter through time, or when the actual direction of the variable changes.

For the linear trend model, the equation that is estimated is of the form:

$$\mathbf{Y_t = a_0 + a_1 t + e}$$

where: **Y** is the variable of interest, **t** denotes the time period (and the time trend when used as an explanatory variable), $\mathbf{a_0}$ and $\mathbf{a_1}$ are coefficients to be estimated, and **e** is the equation error, which represents all of the factors other than the time trend that influence **Y**. The coefficient of the time trend, $\mathbf{a_1}$, gives the estimate of the constant rate of change per period in **Y**.[8] If this coefficient is statistically significant, the per-period rate of change in **Y** is judged to be significantly different from 0. In what follows, the dependent variable, **Y**, is employment for a given nurse category.

In economics, the time trend variable is referred to as the secular, or long-term, trend in **Y**. While it is an important feature of the nurse labor market to estimate, there are a set of other factors at work that are also very meaningful. Among those are the wages of nurses,

[8] Technically, $\mathbf{a_1}$ denotes the change in the mean of **Y** per time period.

working conditions, and labor market conditions, several of which are policy variables. Using trend analysis, it is not possible to ascertain the impact of any of these on nurse employment. Using econometric analysis, however, it is possible to explore the individual effects of the secular trend and the other variables of interest, such as real wages, that have critical policy importance.

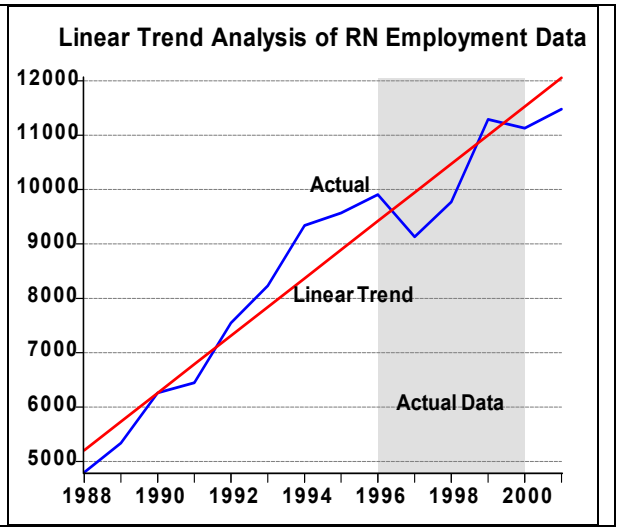
RESULTS FOR RN's

The data from 1988 – 2000 on RN employment were used in a trend analysis, which produced the following estimated equation results (with t-statistics in parenthesis):

$$\hat{Y}_t = \underset{(13.732)}{4{,}679.355} + \underset{(12.270)}{526.762}t \qquad R^2 = 0.932 \quad N = 13 \quad \hat{\sigma}_e = 579.177 \quad \overline{Y} = 8{,}366.7$$

The estimated change in mean RN employment per year is 526.762, which is statistically significant. Overall, the linear trend equation is able to account for 93.2 percent of the observed variation in RN employment based on the equation $\mathbf{R^2}$ statistic. The standard error of this equation, 579.177 is 6.92 percent of the mean for employment, 8,366.7. In terms of the typical measures of equation performance, this equation is satisfactory.

	Actual	Predicted	Error
1988	4,800.1	5,206.1	-406.0
1989	5,338.7	5,732.9	-394.1
1990	6,262.1	6,259.6	2.5
1991	6,449.0	6,786.4	-337.4
1992	7,548.3	7,313.2	235.1
1993	8,229.8	7,839.9	389.9
1994	9,340.0	8,366.7	973.4
1995	9,570.9	8,893.5	677.4
1996	9,908.0	9,420.2	487.8
1997	9,130.0	9,947.0	-817.0
1998	9,770.0	10,473.7	-703.7
1999	11,290.0	11,000.5	289.5
2000	11,130.0	11,527.3	-397.3



The table above gives actual and predicted values of RN employment, along with the estimation errors for each year. As the graph shows, during the period for which actual data exist, 1996 – 2000, the errors resulting from the predictions of trend line are seldom small. This does not necessarily indicate that trend analysis has no value, however.

According to this equation, the estimated annual rate of change in RN employment is 526.762, irrespective of the level of the real wage, its change, the state's unemployment rate, changes in health services employment in Rhode Island, or any other factor. This estimated trend line was used to predict actual 2001 RN employment. While actual RN employment rose slightly from 11,130 in 2000 to 11,480 in 2001, the trend line's

prediction was employment of 12,054, a 5 percent forecast error. When the trend line is estimated only through 1998, its prediction for 2001 rises to 12,140, a 5.75 percent error.

These two features highlight the inherent limitations of linear trend analysis: it presupposes a constant rate of change, even when the levels of other "relevant" variables are changing; and its forecasts tend to be farther off track the more distant are the time periods for which predictions are generated.

Trend analysis can be undertaken where the constant rate of change is not relevant. This entails "higher order" trend analysis. The variant considered here is quadratic trend analysis. The equation that is estimated is of the form:

$$\mathbf{Y_t = a_0 + a_1 t + a_2 t^2 + e}$$

where the variables are defined as in the linear-trend model, except that the additional variable $\mathbf{t^2}$, the square of the time trend, is included. In this model, the rate of change in the mean of **Y** per time period is given by the expression:

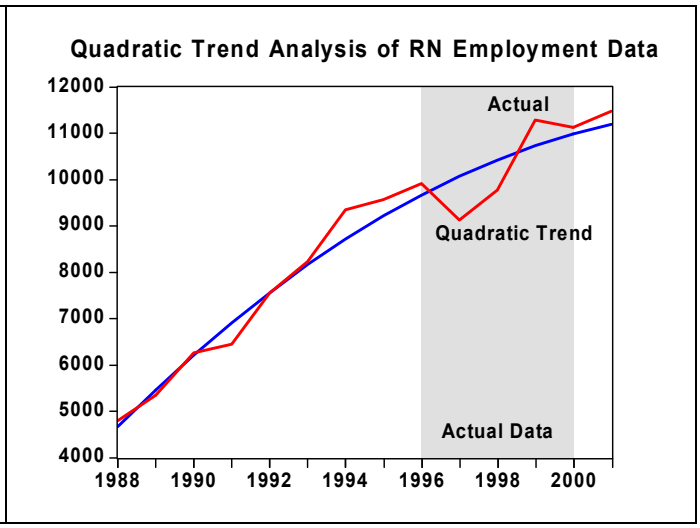
$$\mathbf{a_1 + 2a_2 t = f(t)}$$

which, as noted above, is a function of time. The rate of change in **Y** thus depends upon the time period considered. According to this variant of trend analysis, an upward sloping estimated trend line, such as the one for RN's in Rhode Island, can either become steeper (i.e., increasing at an increasing rate), flatter (increase at a decreasing rate), or reverse direction entirely and begin to decline after some point is reached ($\mathbf{a_1}$ and $\mathbf{a_2}$ differ in sign). The estimated quadratic trend equation for RN's in Rhode Island was estimated. The results are given below:

$$\hat{Y}_t = \underset{(7.833)}{3{,}819.489} + \underset{(5.436)}{870.709}t - \underset{(-2.207)}{24.568}t^2 \qquad R^2 = 0.954 \quad N = 13 \quad \hat{\sigma}_e = 498.154 \quad \overline{Y} = 8{,}366.7$$

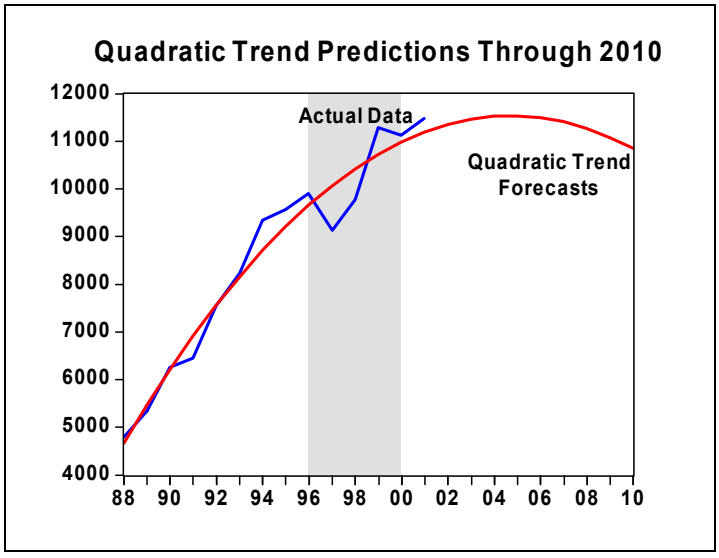
This equation provides a better overall fit to the RN employment data than did the linear trend model, as the equation standard error is lower (498.154 vs. 579.177), and its coefficient of determination is higher ($\mathbf{R^2}$ of 0.954 vs. 0.932). Both time trend coefficients are (individually) statistically significant at the 5 percent level and above, and the entire time trend in this equation is also statistically significant (i.e., the joint significance of both $\boldsymbol{t}$ and $\boldsymbol{t^2}$ via an F-test). The table below provides the actual and predicted values from this equation, along with the error for each year.

QUADRATIC TREND MODEL			
	Actual	Predicted	Error
1988	4,800.1	4,665.6	134.5
1989	5,338.7	5,462.6	-123.9
1990	6,262.1	6,210.5	51.6
1991	6,449.0	6,909.2	-460.2
1992	7,548.3	7,558.8	-10.6
1993	8,229.8	8,159.3	70.5
1994	9,340.0	8,710.6	629.4
1995	9,570.9	9,212.8	358.1
1996	9,908.0	9,665.9	242.1
1997	9,130.0	10,069.8	-939.8
1998	9,770.0	10,424.6	-654.6
1999	11,290.0	10,730.3	559.7
2000	11,130.0	10,986.8	143.2



From the graph on the right, the improved fit of this equation over the linear trend model is visually apparent. This trend line bends through time, indicating that predicted RN employment is increasing at a decreasing rate.

As a further test of this model, the quadratic equation was estimated only through the year 2000 then used to predict the actual 2001 value. Its prediction for 2001, 11,194, is fairly close to the actual value of 11,480, certainly an improvement over the linear trend model's prediction of 12,054. While this looks very promising as a predictive tool, keep in mind that this specific equation's results will continue to indicate that RN employment is increasing at a decreasing rate. Eventually, predicted employment will drop. This is apparent in the graph below, where the estimated equation was used to generate forecasts of RN employment through the year 2010.



Predicted RN employment peaks in the year 2005, at 11,532, and declines thereafter, falling to a value of 10,850 in the year 2010, roughly equivalent to the actual value for

RN employment in the year 2000. According to the predictions of this equation, actual RN employment in Rhode Island is already at its peak through the end of this decade. And, as was true for the linear trend model, *this result pertains irrespective of changes in any policy variables*. So the limitation of trend models, whether linear or of higher order, such as the quadratic trend model here, is that they are blind to current or new information relevant to the variable under study. This is perhaps their greatest limitation, in spite of the relative ease with which they can be estimated.

Clearly, trend models should not be used as the basis for policy concerning labor markets in the future. Models that embody current and likely future information, especially information pertaining to actual or likely future values of policy variables, are needed. Fortunately, there is a way to bridge the gap. This is where econometric analysis comes in.

The errors from trend equations contain important information. These errors, which represent the difference between actual and predicted values, reflect the combined influence of *all factors other than the time(secular)trend* on the employment of RN's. Among the list of potentially influential variables are wages and labor market conditions, some of which are policy variables. Using any variant of trend analysis, it is not possible to ascertain the individual impact of any of these on RN employment. Using econometric analysis, however, it is possible to explore the individual effects of the secular trend and the other variables (such as wages) that have critical policy importance.

To establish the potential relevance of these other variables, the errors from the linear trend analysis reported earlier were used as the *dependent* variable in an equation whose explanatory variables were last year's unemployment rate in Rhode Island, health service employment in Rhode Island as a proportion of Miscellaneous Service employment, and CNA employment as a proportion of health service employment. This equation was able to account for 59 percent of the observed variability of the errors (based on the $\mathbf{R^2}$ statistic), with all of the explanatory variables attaining statistical significance at the 10 percent level and above. While this is hardly a "model" equation, these results point to the likelihood that there is much room for empirical improvement over linear trend analysis.

Real wages for RN's were not statistically significant in the equation with the error terms as dependent variable. This does not indicate that the finding earlier, of relatively slow growth in real wages for this nurse category in Rhode Island, is not related to observed RN shortages. Instead, these results point to the possibility that real wages are related to the secular trend in RN employment (whether it is linear or not). To illustrate how multiple regression methods used in econometrics can improve upon linear trend analysis,[9] let us assume that the secular trend in RN employment is a function only of real wages for RN's. Given the results for the linear trend error equation above, we then have the following:

[9] The same is also true for quadratic trend analysis and orders of trend terms higher than this. Linear is used for discussion purposes since it's elements are fairly easy to grasp and this is the model most persons are familiar with.

RN Employment = f (secular trend, error term)

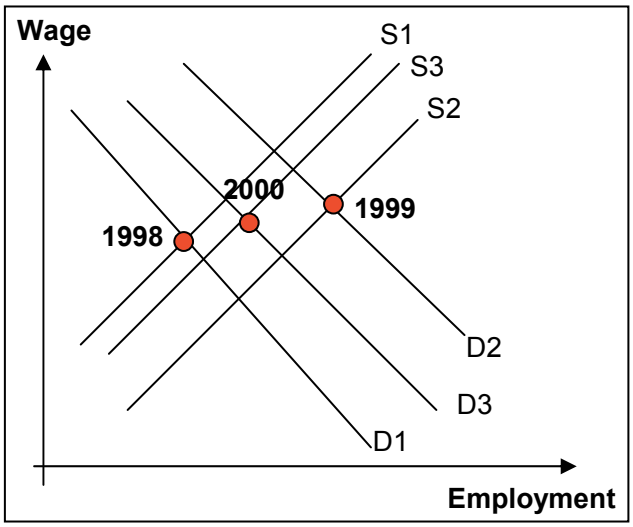
which implies that:

RN Employment = f (RI Unemployment Rate last year, health service employment in RI as a proportion of Service employment, CNA employment as a percentage of health service employment, real wages for RN's, an error for other factors not considered)

This last equation is a multiple regression model. It can provide us with estimates of the *individual* effect of each of the explanatory factors in parenthesis above. More importantly, these individual effects are partial rates of change: they pertain to a change in a single variable, given the value of every other explanatory variable in the equation. This is an important and powerful feature of multiple regression analysis. Furthermore, when highly relevant factors are not included among the variables in such an equation, their influence tends to spread to the included variables. The result is what statistics refers to as bias. So, well specified multiple regression equations for employment hold the promise of greatly extending the knowledge gained from linear trend analysis. Instead of predicting a constant rate of change no matter how much relevant policy variables are changing, multiple regression analysis can generate predictions based on specific changes in policy variables, which can either be forecasts, when realistic values for policy variables are considered, or simulations when hypothetical possible values are used.

DATA ISSUES AND PROBLEMS

The employment data that are available can be analyzed econometrically, but, the task is not as simple as it might first appear to be. Based on economic theory, employment, or the equilibrium quantity of any variable, is determined by the interaction of demand and supply. The graph below illustrates the nature of available data.



Observed RN employment for each year pertains to neither a single demand curve nor a supply curve. Instead, it emerges from the intersection of a series of demand and supply curves that are changing over time. This problem is referred to as the identification problem in econometrics. It arises because wage rates are not determined in isolation of employment. Instead, the values of wages and employment each year are jointly (simultaneously) determined.[10] Dealing with this requires not only adding variables to a single equation (as was done for multiple regression), but adding additional *equations* so that there is a separate equation for each variable that is part of the simultaneous relationship. In the present context, to correctly model statewide employment of nurses, it will be necessary to generate *separate models* for the supply and demand of nurses in each nurse category. Each of these equations will contain the wage rate as an explanatory variable and other variables that distinguish it from the other equation under consideration.

- *The goal of such modeling is to isolate (identify) the separate curves to ascertain how much wages affect demand and supply individually.*

ANOTHER TIME SERIES ESTIMATION METHOD

Acknowledging that employment and real wages need to be *jointly* modeled, a framework exists within this can be accomplished without the need for all the data gathering that is required in a full (simultaneous equation) econometric model. This time series method is referred to as Vector Autoregression Analysis (VAR). In this framework, the value of either employment or real wages is viewing as being determined by its own past values (i.e., lags of that variable) and past values of the other jointly determined variable. The employment equation, for example, might be of the following form:

$$\mathbf{E_t = a_0 + a_1E_{t-1} + a_2RW_{t-1} + e_t}$$

where **E** represents employment, **RW** is the real wage, $\mathbf{a_0}$ through $\mathbf{a_2}$ are coefficients to be estimated, and **e** is the error term. In this equation, it is assumed that one lag is relevant in modeling employment.[11] If two lags pertain, this model becomes:

$$\mathbf{E_t = a_0 + a_1E_{t-1} + a_2E_{t-2} + a_3RW_{t-1} + a_4RW_{t-2} + e_t}$$

For RN employment in Rhode Island, the first of these models fits the data better. Using this with the data for 1988 through 2000 and forecasting the 2001 value, the result is very good: the predicted value, 11,318, differs by only 1.4 percent from the actual value in 2001, 11,480. While this technique is, in many ways, preferable to either of the variants of trend analysis considered earlier, it too has limitations. Like trend analysis, this is "blind" to current information which may be relevant in terms of actual or contemplated

[10] In econometric terminology, both of these variables are endogenous – their values are determined from a system of equations.

[11] In practice, the appropriate lag length must be determined empirically. Typically, either the Akaike Information Criterion or the Schwartz Information Criterion is used to make this determination.

policy. As a result, its usefulness in terms of evaluating policy alternatives is somewhat limited. It is a worthwhile tool to use in producing short-term forecasts, however.

ANOTHER ATTEMPT TO "FILL IN THE GAPS"

Efforts to augment existing data were not limited to the "macro" labor market. During the time of this grant, questions were formulated and submitted to the Rhode Island Department of Health that were included on the CNA license renewal form. These "micro" focus questions, attempted to ascertain information on the following:

- Whether currently employed;
- Employed in-state, out-of-state, or both;
- Part-time or full-time employment; and
- Employment setting.

A copy of the questions that were included on the license renewal form is given below.

Employment Data:

(Optional)

Are you currently employed as a Nursing Assistant? ☐ Yes No

Setting: (check all that apply)	Numbers of Hours Worked per Typical Week:	Setting: (check all that apply)	Number of Hours Worked per Typical Week:
☐ Assisted Living Residence	☐☐☐	☐ Home Care Provider	
☐ Hospice Center	☐☐☐	☐ Hospital	
☐ Nursing Home	☐☐☐	☐ Nursing Service Agency	
☐ Outpatient Rehabilitation Center	☐☐☐	☐ Organized Ambulatory Care Facility	
☐ Kidney Disease Treatment Center	☐☐☐	☐ Physician Office Setting	
☐ Rural Health Clinic	☐☐☐	☐ School Based Health Center	
☐ Freestanding Emergency Care Facility	☐☐☐	☐ Freestanding Ambulatory Surgery Center	

When analyzed, this information can form the basis of an extensive and ongoing database. Unfortunately, we have not yet been granted access to the responses to these questions, nor have they been included in subsequent license renewal forms. It is our belief that information such as this can potentially provide very important insights that will be useful in formulating policy analysis. Hopefully, information such as this can be regularly gathered and analyzed, creating an ongoing database of great usefulness.