Irish Industrial Wages: An Econometric Analysis

Edward J. O'Brien - Junior Sophister

With pay agreements firmly back on the national agenda, Edward O'Brien's topical econometric analysis aims to identify potential factors which determine the level of average industrial wages in Ireland and, in particular, might account for strong county to county variation. Rates of urbanisation and unemployment, disposable income and whether a county's industries are favoured with access to a seaport are tested for their influence on the wage level.

Introduction

The concept of the average industrial wage is one with which many of us are familiar. The average industrial wage is the average annual wage paid to those in the Industrial sector. The Industrial sector, as defined by the Central Statistics Office, includes the manufacturing, mining, and quarrying, and the electricity, gas and water sectors only.

The average industrial wage in Ireland varies considerably across the twenty-six Counties. The average value for the entire country is £14,461.41. However, extremes can be found in Counties Dublin and Donegal, where wages are £18,334.41 and £10,853.24 respectively.

This essay aims to explain these variations, using a regression analysis. I shall begin by formulating an econometric model, and discussing the variables used therein. I shall then estimate the model and analyse the results in some detail. Finally, I shall conclude the essay, with a critical evaluation of the model.

The Econometric Model

The multiple regression model I shall use is as follows:

$$\text{Average industrial wage} = \beta_0 + \beta_1 (\text{ \% Urban }) + \beta_2 (\text{ \% Unemployment }) + \beta_3 (\text{ Port }) + \beta_4 (\text{ Disposable Income }) + \mu$$

As can be seen from the model, I have chosen four variables to explain the variations in the average industrial wage. I shall explain each in turn.

Dependent Variable Y

This essay aims to explain variations in the average industrial wage across the counties of Ireland. Therefore, the dependent variable is this wage rate. Twenty-six wages are taken, one for each county. This data was obtained from the *Census of Industrial Production, 1996.*

First Independent Variable X_1

The first explanatory variable is the percentage urbanisation of a county. This value signifies what proportion of the county's population live in an urban environment. The CSO defines *'urban'* as a town with 1,500 people or more. Any settlement with less than 1,500 people is considered to be rural.

Industrial Location Theory suggests that urbanisation is relevant to this analysis. Industry requires a labour force. Therefore it tends to locate in urban areas, where a labour force is available. Areas, or counties with higher levels of Industrialisation are likely to benefit from higher wage rates.

Second Independent Variable X_2

The second explanatory variable is the percentage unemployment of a county. These figures also relate to 1996, and refer to the total number of unemployed, excluding first time job seekers, as a percentage of the total county population.

Theory suggests, that where unemployment is higher, wages should be lower. The labour market is not in equilibrium. An excess of labour supply exists. Therefore, wages can fall without reducing the supply of ready labour.

Dummy Variable X_3

The third explanatory variable is port access. This is a simple dummy variable. Transportation is a major consideration for Industry. Both inputs and outputs must have easy access to production facilities and markets. Port sites are particularly beneficial, particularly in the manufacturing, mining and quarrying industries. In these locations, Industry will be more competitive and successful. As a result, higher wages may be paid.

The fourth, and final explanatory variable is the Index of Disposable Income. This also relates to the year 1996. It is intuitively appealing to suggest that wage rates will reflect the costs of living. If the cost of living differs across Counties, then one would expect wage rates to differ. The index reflects this fact.

Estimation

I shall use the method of ordinary least squares in this analysis. This method produces the line of best fit, upon which my econometric model is based:

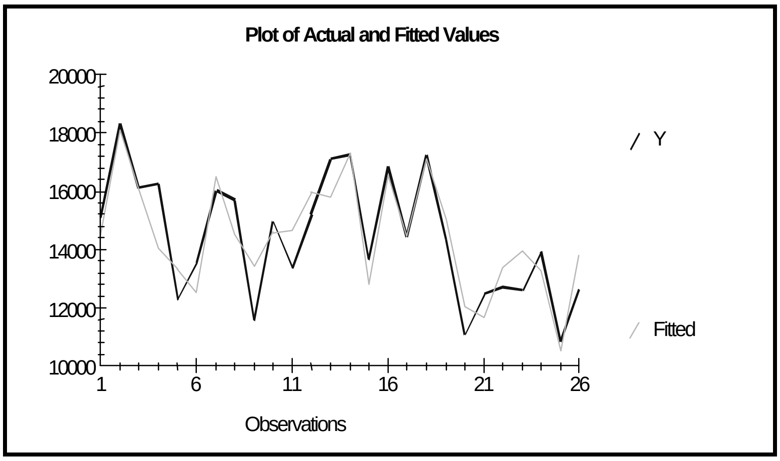
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \mu$$

where μ is the stochastic disturbance term.

The regression is: $Y_i = 22{,}103.0 + 94.72X_1 - 877.06X_2 + 1{,}165.8X_3 - 77.14X_4$

Correlation

The Correlation coefficient for this regression was found to be 78%, or $R^2 = 0.78$. This value for R^2 is quite high. R-bar-squared was found to be 0.74. The correlation can be seen below in *Figure 1*, the plot of actual and fitted values of Y.



Analysis:

Multiple Regression

Table 1, summarises the regression results

Independent Variable	*Coefficient*	*t - Statistic*	*Probability*
Constant C	22,103.0	5.8802	0.000
X_1	94.7202	5.6250	0.000
X_2	-877.0633	-3.3184	0.003
X_3	1,165.8	2.1910	0.040
X_4	-77.1351	-2.1126	0.047

It is worth considering the above results, and comparing them with prior expectations. In the case of urbanisation, X_1, it was expected that the coefficient would be large and positive. In fact, it is positive, but perhaps not as large as one might have expected. X_2 also, is as anticipated. It was expected that unemployment would have a large and negative coefficient. This is in fact so. port access, X_3, also returns expected values. It is positive and large. The only unexpected result was disposable income, X_4. The value is large, as expected, but it is negative, when a positive result was expected. This is difficult to explain.

Single Regressions:

Below, in the following tables each variable is regressed individually on Y.

Table 2.

Independent Variable	*Coefficient*	*t - Statistic*	*Probability*
Constant C	11,223.2	18.4385	0.000
X_1	83.3070	5.9158	0.000

$R^2 = 0.59319$

Table 3

Independent Variable	*Coefficient*	*t - Statistic*	*Probability*
Constant C	14,253.4	5.8832	0.000
X_2	40.7333	0.087152	0.931

$R^2 = 0.0003164$

Table 4

Independent Variable	*Coefficient*	*t - Statistic*	*Probability*
Constant C	13,404.3	33.3028	0.000
X_3	2748.4	4.2348	0.000

R^2 = 0.42767

Table 5

Independent Variable	*Coefficient*	*t - Statistic*	*Probability*
Constant C	7,350.9	1.3325	0.195
X_4	75.2003	1.2924	0.209

R^2 = 0.065066

As can be seen from the above *Tables,* the single regressions give some interesting results. Both X_1 and X_3 behave as expected, with positive and large coefficients. The values of R^2 are 0.59 and 0.42 respectively. Also, the t-statistics for these variables return a zero probability, and hence are significant at the 1% level.

However, the variables X_2 and X_4 return results that are difficult to explain. The size and sign of the coefficient of X_2 are unremarkable. However, the constant term is large, and by comparison the coefficient is small. The Unemployment data has a very small range, between approximately 3% and 8%. This results in a regression equation that is almost a straight line. The probability that $\beta_2 = 0$ is 93%. This is strengthened further by the most conclusive R^2 value found so far, 0.003164. That is to say, Y and X_2 are completely unrelated.

A similar pattern is found with X_4. Although the data for X_4 is not of such a small range as X_2, the regression equation again is almost a straight line. The probability is 20.9%, so one would still accept the null hypothesis, i.e. $\beta_4 = 0$. Also, the value of R^2 is very low, 0.065066. Again, this suggests almost no relation between Y and X_4.

The above problems are difficult to interpret. The multiple regression is good. It has a high R^2 value and a 5% level of significance. This would imply that the model is good. However, the individual regressions tell a different story. The most probable explanation for such a discrepancy is multicollinearity between the X variables.

To examine if this was actually the case, regressions were run between each of the X variables. Only X_1 was significant, having R^2 values of 0.36, 0.58, and 0.48

respectively, when regressed upon X_2, X_3, and X_4. This confirms that at least some of the above problems are due to multicollinearity.

Statistics

T – Statistic
The t-statistic is a measure of the ratio of the estimate to the standard error. The t-statistic allows the hypothesis that the coefficients equal zero to be tested. That is:

$$H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \text{ and } \beta_4 = 0.$$
$$H_1: \beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \text{ and } \beta_4 \neq 0.$$

In the multiple regression, X_1 and X_2 were found to be significant at the 10%, 5% and even the 1% levels. X_3 and X_4 were significant at the 10% and 5% levels. Therefore the null hypothesis H_0 can be rejected as the coefficients do not equal zero.

However, the single regressions gave somewhat different results. Both X_1 and X_3 were significant to the 1% level. Therefore, the null hypotheses can again be rejected. However, both X_2 and X_4 are not even significant at the 20% level. Therefore, the null hypothesis must be accepted, i.e. H_0: $\beta_2 = 0$; $\beta_3 = 0$.

F – Statistic

Whereas the t-statistic evaluates the significance of individual variables, the F-statistic evaluates the combined significance of all variables. The F-statistic for this model is $F_{4,21}$ = 18.86777. At this value, there is negligilbe probability. This confirms the results from the t-statistic. The null hypothesis, H_0: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$, and $\beta_4 = 0$, can be rejected. The coefficients are not all zero. That is to say, the model has *some* explanatory power.

Forecasting

In order to ascertain the predictive power of the model, the last four observations were omitted from the regression. One would expect that a reasonable model should have some predictive capabilities. The results can be seen below, in *Table 6.*

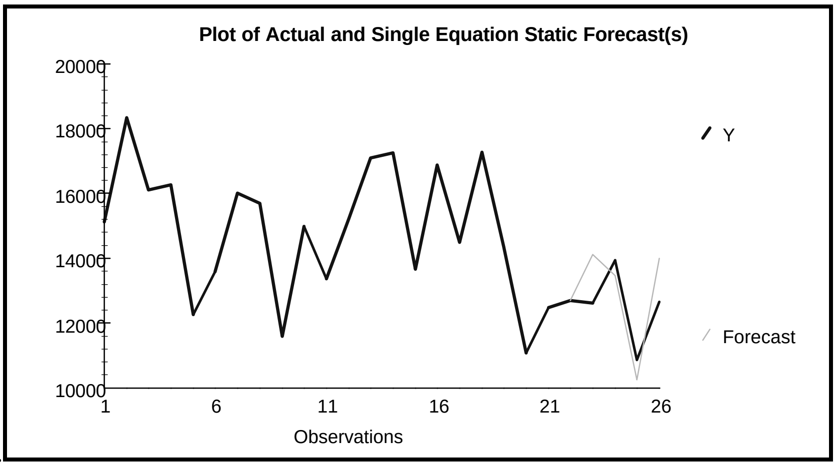
Table 6.

County	*Actual*	*Predicted*	*Difference*
Sligo	£12,619.1	£14,127.1	-£1,507.9
Cavan	£13,941.9	£13,469.8	£472.04
Donegal	£10,853.2	£10,249.0	£604.21
Monaghan	£12,663.3	£14,008.2	-£1,344.9

As can be seen from *Table 6*, the model has some predictive power, yet is not very accurate. The mean prediction error is -£444.14. Although this value seems small, one must remember that this is an arithmetic average of both positive and negative difference values. A better statistic is the Mean Sum Absolute Prediction Error. This avoids the above problem by taking absolute values of the differences. It value is £982.26. So the predictions, on average, are inaccurate by almost £1,000. This equates to an error of between 7.0% and 9.0%, approximately, in each of the four observations.

Below, in *Figure 2*, the graphical representation of the forecasting ability is shown. Here it can be seen that while the model does follow actual trends, it fails to be very accurate in these trends.

Figure 2:



Conclusion

The econometric model used can be deemed to be a success. It yields satisfactory values for R^2, t-statistic, and F-Statistics. Using urbanisation, unemployment, port access, and an index of disposable income, the model satisfactorily explained variations in the average industrial wage. A value of $R^2 = 0.78$, with a 1% significance from the F-Test is strong evidence of definite correlation. This was further borne out by the forecasting ability of the model.

However, some problems were noted, specifically in the single regressions.

- Some multicollinearity is evident. This was particularly obvious with X_1. However, this makes perfect sense. Unemployment tends to be higher in urban areas. In Ireland, all major ports are found in urban centres. The level of disposable income is generally higher in urban areas. These factors account for the presence of multicollinearity.

- Omitted variables are sure to be a problem. The average industrial wage is complicated by a wide variety of factors. Many of these externalities have been omitted from this model. Indeed, many of them remain unknown and would require additional research to be understood.

- Data problems also exist. The average industrial wage is an arithmetic mean, and suffers from being just an *average* value. Outlying data and the necessary arbitrariness of definitions may cause distortion. Finally, the Index of Disposable Income is based on regions within Ireland, not counties.

However, despite the above reservations, this econometric model has provided many useful insights taking into account the above difficulties.

Appendix

Data Sources

- The average industrial wage is taken from the Census of Industrial Production, 1996.
- The percentage urbanisation is taken from the Census of Population, 1996.
- The percentage unemployment rate is also taken from the Census of Population, 1996.
- The port access data is taken from ‘Facts About Ireland’.

- The Index of Disposable Income is taken from the Household Budget Survey, 1994 -95.

Reliability

- Almost all data is taken from Central Statistics Office (CSO) publications. Therefore one can assume it is the most accurate and reliable data available.
- The only data not taken from the CSO is that on port access. This was taken from a government publication. The nature of the data also rules out inaccuracy.

Bibliography

Gujarati, D. (1995) *Basic Econometrics*. McGraw-Hill: Singapore.

Kennedy, P. (1992) *A Guide to Econometrics*. Blackwell: Oxford.

Drudy, P. J. and M. Punch (1999) *The "Regional Problem", Urban Disadvantage and Development*. Trinity College: Dublin

Department of Foreign Affairs (1995) *Facts About Ireland*.

Central Statistics Office (1996) *Census of Industrial Production*.

Central Statistics Office (1996) *Census of Population*.

Central Statistics Office (1995) *Household Budget Survey*.

Acknowledgements
Professor. P.J. Drudy, Department of Economics, Trinity College, Dublin.
Margaret Power, Central Statistics Office, Cork.