An Econometric Study of the Labor Demand in Hong Kong

LAU Ho Ming* City University of Hong Kong, Hong Kong, China, <u>hominglau6@cityu.edu.hk</u>

Abstract

Based on one of the feature articles entitled "The Situation of the Four Key Industries in the Hong Kong Economy in 2010" that appeared in the February 2012 issue of the Hong Kong Monthly Digest of Statistics, 'Trading and Logistics' is one of the Four Key Industries in the Hong Kong Economy. The value added of this sector is the largest in the Hong Kong Economy and its share in GDP in 2010 was 25.3%. In view of its importance in the labor market by absorbing 22.4% of the total employment in Hong Kong in 2010, the purpose of this study is to investigate factors that affect behind this sector. To give insights into the relationship between the labor demand and factors, our objective is to formulate an effective econometric model to estimate and predict the labor demand in the trading industry in Hong Kong. By utilizing with Cobb-Douglas production function and firm's minimization principle, we derive the conditional labor demand function which is dependent on the wage rate, unit cost of capital and value added and formulate the econometric model. By examining and checking the historical data, we found that factors (such as Labor Wage) and major events (such as Financial Crisis) affect the labor demand. We will include dummy variables in the econometric model to model major events and improve prediction. This presentation will describe the econometric model used in this study. Our findings will advance our understanding of the labor demand in Hong Kong's trading industry.

Keywords: Cobb-Douglas production function, Conditional labor demand function, Cost minimization problem and Double-log regression model

1. Introduction

The government, politicians, economists and firms are always interested in the impact on labor demand when some changes have been made in the market, such as business and economic cycle fluctuations, minimum wage policy, technology shocks and unionization. During the past century, different researches have been done to investigate the in-depth relationship between labor demand and several factors. However, Hong Kong's labor demand was seldom to be investigated by others. In view of this issue, Hong Kong's labor demand will be investigated in the following study. To provide the in-depth investigation in the specific industry in Hong Kong, the trading industry has been selected as being the largest and most important industry in the Hong Kong economy, whose share in GDP and the labor market are the largest in Hong Kong. In the following sections, the econometric model will be identified from the theoretical model derived by Cobb-Douglas production function and firm's minimization problem to estimate and predict the labor demand in Hong Kong's trading industry.

2. Theoretical Model

In view of Varian (1978), the conditional labor demand function in a firm can be derived from minimizing the total cost of producing Y units of output as follows:

Assume that two factor inputs, the capital and labor, are used to produce *Y* units of output in a firm. The relationship between output level and factor inputs can be described in a production function. From Durand (1937), he suggested the general Cobb-Douglas production function should be

$$Y = AL^{\alpha}K^{\beta} \tag{1}$$

where Y is output, A is technology level, L is labor input, K is capital input and α and β are constants.

For each choice of L and K, there will be some choices that minimize the total cost of producing Y units of output. Let w be the wage rate and r be the unit cost of capital as the factor input prices of L and K. Considering the cost minimization problem subject to the fixed output level and factor input prices, the total cost function, denoted as c, is

$$c(w,r) = \min_{L,K} wL + rK$$

$$s.t. Y = AL^{\alpha}K^{\beta}$$
(2)

The conditional demand function for labor in a firm is

$$L(w,r,Y) = A^{-1/\alpha+\beta} \frac{\alpha r^{\beta/\alpha+\beta}}{\beta w} Y^{1/\alpha+\beta}$$
(3)

From Hamermesh (1986), the conditional labor demand function (3) which is dependent on the wage rate, unit cost of capital and output level can be used to formulate the econometric model for the sake of gaining insights into the relationship between the labor demand and several factors.

3. Econometric Model

Based on the theoretical model (3) derived in Section 2, we use the natural logarithm transformation on (3) to get the following double-log functional form:

$$\ln L = -\frac{1}{\alpha+\beta}\ln A + \frac{\beta}{\alpha+\beta}(\ln\alpha - \ln\beta) - \frac{\beta}{\alpha+\beta}\ln w + \frac{\beta}{\alpha+\beta}\ln r + \frac{1}{\alpha+\beta}\ln Y.$$
(4)

The double-log functional form in (4) will be used to formulate the econometric model for time series data in the whole study as follows:

$$\ln L_t = \beta_0 + \beta_1 lnw_t + \beta_2 lnr_t + \beta_3 lnY_t + \varepsilon_t,$$
(5)
where $\beta_0 = -\frac{1}{\alpha+\beta} lnA + \frac{\beta}{\alpha+\beta} (ln\alpha - ln\beta), \beta_1 = -\frac{\beta}{\alpha+\beta}, \beta_2 = \frac{\beta}{\alpha+\beta}, \beta_3 = \frac{1}{\alpha+\beta} \text{ and } \varepsilon_t \text{ is the error term.}$

There are two main reasons to transform (3) into the double-log functional form. Firstly, this functional form is particularly useful for estimation purpose as providing a linear function so that OLS method can be applied for the estimation. Secondly, this functional form allows the associated coefficients to have elasticity interpretations. For example, the coefficient (β_1) associated with the variable (lnw_t) means that one percent change in the wage rate would, on average, result in β_1 percent change in the labor demand, holding other factors constant.

Generalizing the concept to an industry to assume that an industry is a firm, the econometric model (5) can be used to model the labor demand in Hong Kong's trading industry and will be used in the whole study.

4. Data used in the econometric model

To gain the preliminary insight into the labor demand, wage rate, unit cost of capital and output level, annual data from 1980 to 2011 are retrieved from the website of Hong Kong Census and Statistics Department and presented, respectively, in Table 1 and Figure 1 as follows:



where L is number of employees, w is compensation of employees divided by number of persons engaged, r is operating expenses and Y is value added. The variables on w, r and Y are in real terms with 2010 as the base year. Note that due to the unavailability of data in the unit cost of capital, the operating expenses is regarded as a proxy variable for the unit cost of capital to avoid the omit variable bias during OLS estimation.

From Table 1, it is obvious that each time series exhibits an increasing trend. A reason behind this phenomenon is probably the economic growth and prosperity. To investigate the pattern of labor demand in detail, from Figure 1, the significant and sudden drops were found in 1997, 1998, 2001, 2002 and 2008. To figure out the possible reasons behind these drops, we check the historical events. The outbreak of Asian financial crisis in 1997 had adversely affected the economy in Hong Kong in 1997 and 1998. In 2001, the September 11 attacks in US had decreased the trading activities as the strict restriction on airspace was imposed and had deteriorated the US and global economies. Moreover, based on one of the feature articles entitled "The Import/Export, Wholesale and Retail Trades, and Accommodation and Food Services Sectors in Hong Kong" that appeared in the May 2011 issue of the Hong Kong Monthly Digest of Statistics, the outbreak of global financial tsunami in 2008 brought a negative impact on the trading industry. All major events directly affected the labor demand in Hong Kong as the demand and supply in outputs were significantly decreased during these periods. Therefore, these events had caused the significant and sudden drops in 1997, 1998, 2001, 2002 and 2008.

To estimate the econometric model (5) defined in Section 3, the data in Table 1 will be indexed to represent each variable in the econometric model as the annual data are recorded in terms of the dollars. In order to gain a precise real value, indexing the data based on 1980 is essential to eliminate the adverse effect of large scale in the data. The representation of each variable in the econometric model (5) is as follows:

 L_t Index of number of employees per year (Base = 1980)

 w_t Index of compensation of employees divided by number of persons engaged in real terms per year (Base = 1980)

 r_t Index of operating expenses in real terms per year (Base = 1980)

 Y_t Index of value added in real terms per year (Base = 1980)

In order to better capture the major historical drops which had affected the demand for labor (L), the following three dummy variables are introduced into (5).

| $D_{1,t} = \Big\{$ | 1 ∀ t = 1997, 1998 0 otherwise | Model the Asian financial crisis in 1997 and 1998 |
|-------------------------------|---|--|
| $D_{2,t} = \Big\{$ | 1 ∀ <i>t</i> = 2001,2002 0 otherwise | Model the September 11 attacks in 2001 and the aftermath in 2002 |
| $\boldsymbol{D}_{3,t}=\Big\{$ | 1 ∀ t = 2008 0 otherwise | Model the Global financial tsunami in 2008 |

Note that two data points in 2010 and 2011 will not be used in the model estimation in Section 5 for the validation purpose.

5. OLS Estimation and Empirical Result

As a starting point, the econometric model without any dummy variables (5) is estimated by the OLS method as follows:

$$\begin{aligned} & \ln \bar{L}_t = 3.23759 - 0.68693 ln w_t + 0.41374 ln r_t + 0.57998 ln Y_t + \hat{\varepsilon}_t \\ & (t \ ratios) \ (3.84) \\ & (-2.64) \\ & (3.45) \\ F \ ratio = 196.97, \ R^2 = 0.9579, \ Adj. \ R^2 = 0.9530, \ DW = 0.528, \ n = 30 \end{aligned}$$

The result indicates a strong relationship between labor demand and several factors. It seems very attractive if considering the high *F* ratio, t ratios and R^2 solely. However, when modeling with time series data, the most important issues are to test for any spurious regression and serial correlation in the estimated model. From Granger and Newbold (1974), the rule of thumb is suggested that with time series data, if the R^2 is greater that the *DW* statistic, it is suspicious to have a spurious regression. From the result, R^2 (0.9579) is greater than the *DW* statistic (0.528). The apparent relationship seems spurious as the variables in the regression may be non-stationary. Non-stationary time series variables should not be include in a regressions model as two major adverse effects will come out. Firstly, from Granger and Newbold, the spurious regression leads time series to appear related when they are not due to share the common trend, providing the misleading t ratios, F ratio and R^2 eventually. Secondly, from Dickey and Fuller (1979 and 1981), the distribution of the OLS estimator is not normal distributed under the null hypothesis, even in large samples. This leads conventional hypothesis tests and confidence intervals are not valid anymore and become misleading. To test for non-stationary time series data, Dickey and Fuller (1979) have introduced the Dickey Fuller test as follows:

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \alpha_2 t + \varepsilon_t \tag{7}$$

If the null hypothesis $(H_0: \gamma = 0)$ is not rejected, the time series Y_t contains a unit root, hence is non-stationary. To perform the hypothesis test, t-statistic for Y_{t-1} is computed in the usual way but compared with the Dickey-Fuller tables as the conventional hypothesis test is not valid under presenting the non-stationary time series data. The unit root test for each time series variable is shown in Table 2:

Table 2 Unit Root Test for Stationarity in original data: Ho=I(1) vs. H1=I(0) Variables t-statistic

Lt -0.96 wt -2.48 rt -1.47 Yt -1.22

Based on the Dickey-Fuller tables from Fuller (1976), the critical value with 5% significant level is -3.60. It is obvious that all time series data cannot reject the null hypothesis and this provides significant evidence that they are not I(0) and stationary. To deal with this problem, the appropriate route is to transform the variables before running a regression in order to make them stationary. The remedy used here is the differencing transformation:

$$\Delta Y_t = Y_t - Y_{t-1} \tag{8}$$

By adopting differencing transformation (8), the null hypothesis in the unit root test for each time series variable is rejected as shown in Table 3 and thus it is evident that they have achieved stationary and can be used in the regression model. However, note that the number of observation used in the estimation is decreased by 1 from 30 to 29 after the first difference.

Table 3

t-statistic

Unit Root Test for Stationarity in differenced data: Ho=I(1) vs. H1=I(0)

Variables

| Diff Lt | -4.40 |
|---------|-------|
| Diff wt | -6.52 |
| Diff Yt | -4.47 |

After data transformation, the time series data are eventually on the right track and can be used in the following study. In the following part, backing to the objective, the OLS methods can be applied again to estimate the econometric model (5). As mentioned in Section 4, three dummy variables $(D_{1,t}, D_{2,t}, D_{3,t})$ are considered to model the major events which had affected the labor demand in Hong Kong's trading industry. The econometric model (5) with three dummy variables is therefore estimated by the OLS method to improve prediction as follows:

$$\overline{\Delta \ln L_t} = 0.04936 - 0.48936 \Delta \ln w_t + 0.28347 \Delta \ln r_t + 0.05923 \Delta \ln Y_t - 0.05676 D_{1,t} - 0.06518 D_{2,t} - 0.10929 D_{3,t} + \hat{\varepsilon}_t$$
(9)
(t ratios) (4.34) (-1.90) (3.44) (0.54) (-1.57) (-2.20) (-2.75)
F ratio=9 13, R²=0 7136, Adi, R²=0 6354, DW=1 484, n=29

Before any interpretation of the results in the estimated model (9), it seems that the dummy variable which models the Asian financial crisis in 1997 and 1998 ($D_{1,t}$) is not significant under the 5% significant level with 22 degrees of freedom = -2.07. In regard to this insignificance, the dummy variable ($D_{1,t}$) is dropped from the study and the econometric model (5) is now estimated with only two dummy variables ($D_{2,t}$, $D_{3,t}$) as follows:

$$\begin{aligned}
\Delta \widehat{\ln L_t} &= 0.04476 - 0.68882 \Delta lnw_t + 0.29021 \Delta lnr_t + 0.14020 \Delta lnY_t - 0.05430 D_{2,t} - 0.10742 D_{3,t} + \widehat{\varepsilon_t} \\
(t ratios) (3.95) & (-2.98) & (3.42) & (1.42) & (-1.83) & (-2.62) \\
F ratio=9.85, R^2 = 0.6816, Adi, R^2 = 0.6123, DW = 1.325, n = 29
\end{aligned}$$
(10)

The result seems more persuasive then the previously estimated model (6) and (9). As mentioned previously, when modeling with time series data, the tests for any spurious regression and serial correlation in the estimated model are indispensable.

Firstly, to test for any spurious regression, from the result, R^2 (0.6816) is less than the *DW* statistic (1.325). It is evident that the spurious regression is resolved after data transformation under the rule of thumb by Granger and Newbold. At the same time, t ratios and F ratio have restored to the normal level rather than the inflated level when compared with the estimated model (6).

Secondly, serially uncorrelated errors play an important role in OLS method. Without this assumption, two major drawbacks occur when using OLS method. Firstly, OLS estimator is no longer the best linear unbiased estimator (*BLUE*) as it will not have the minimum variance among all other linear unbiased estimators. Secondly, the OLS standard errors and test statistics are no longer valid for the statistical inference, even asymptotically. The presence of serially uncorrelated errors, therefore, is essential in the model. To identify the serial correlation, the Durbin-Watson (DW) statistic based on least squares residuals ($\hat{\varepsilon}_t$) is commonly used. From the result in (10), DW=1.325, however, lies in the inconclusive region which is between $d_L=1.050$ and $d_U=1.841$. From Breusch (1978), *LM test* can be used instead to identify any serial correlation as follows:

$$\hat{\varepsilon}_t = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1} + \alpha_2 \Delta ln w_t + \alpha_3 \Delta ln r_t + \alpha_4 \Delta ln Y_t + \alpha_5 D_{2,t} + \alpha_6 D_{3,t} + v_t$$
(11)

$$LM Statistic = (n-1)R^2 \sim \chi^2_{(1)} \tag{12}$$

By running the auxiliary regression (11) to regress $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}$ and all independent variables and saving the R^2 , *LM Statistic* can be computed as (12) and follows the chi-square distribution with 1 degree of freedom under the null hypothesis of no serial correlation. The *LM Statistic* in the auxiliary regression (11) is $(n-1)R^2=28*0.0625=1.75$ which is much less than the critical value with 5% significant level (3.84). As the null hypothesis is not rejected, there is no evidence that the serial correlation is present in the estimated model (10).

Note that under the Gauss-Markov assumptions, OLS is the best linear unbiased estimator (*BLUE*) and has the usual sampling variances. In view of this issue, before any interpretation of this model, the diagnostic checking for the Gauss-Markov assumptions is indispensable. Otherwise, the interpretation is invalid if any assumption is violated in the estimated model. The residual analysis is widely used in the diagnostic checking to identify whether there is any violation of the Gauss-Markov assumptions. By employing the residual analysis, all assumptions fulfill successfully in the estimated model (10). However, as the most important issues in the time series modeling which are the spurious regression and serial correlation have been discussed preciously, in the interest of preserving space, the detailed discussion of the residual analysis has been omitted but it can be obtained upon request.

To conclude, it is not a coincidence that all assumptions fulfill successfully as the data used in the estimated model (10) has been transformed by using the natural logarithm and first difference, which are the common remedies used when any assumption is violated in the model. After the diagnostic checking, investigating and interpreting the result in the estimated model (10) can be started in the following part.

Interpretation of the Results

As a starting point, checking whether the sign is consistent with the theoretical model is crucial. If there is any wrong sign in the estimated model, the estimated model is useless as the estimated coefficients would not be economically meaningful and the result will thoroughly ruin the whole theory discussed in Sections 2 and 3. From the econometric model (5), the expected sign of the wage elasticity of labor demand is negative. This conforms to the firm's behavior that the higher wage rate will tend to decrease the labor input and use another factor input, which is capital in this model, to substitute for labor and maintain the same output level. In view of the same behavior, the expected sign of the unit cost of capital elasticity of labor demand is positive. For the output level, to produce a higher level in a firm, increasing the factor input is inevitable. Therefore, the expected sign of output level elasticity of labor

demand is positive. For two dummy variables, as they model the major events which brought the negative impacts on the trading industry, the expected sign of them is negative. In the individual perspective, all estimated coefficients in the estimated model (10) have the correct signs as expected.

For the overall significance, the F ratio is large enough to reject the null hypothesis and thus provides significant evidence that the factors in the estimated model (10) are useful to explain the labor demand. At the same time, the high adjusted R^2 indicates the model which has the high goodness of fit to the historical data. In regard to the individual significance, based on the t ratio for each parameter to perform the conventional hypothesis test, except for the output level (Y) and the dummy variable ($D_{2,t}$), other parameters are highly significant as having the high t ratios. However, the output level (Y) will not be dropped from the econometric model (5) due to the theory support and to avoid the omit variable bias during OLS estimation and the dummy variable ($D_{2,t}$) will not be dropped as it just barely insignificant.

To get more insight into the relationship, especially the quantitative relationship, between the labor demand and factors, the coefficient value can be examined. Based on the estimated model (10), it seems that the coefficient of the wage rate (w) whose absolute value is the highest when compared with the unit cost of capital (r) and output level (Y) has the major effect on labor demand in Hong Kong's trading industry. To be quantitative, the coefficient means when the wage rate (w_t) is increased by 1%, the estimated labor demand (\hat{L}_t) is decreased by about 0.69% on average, holding other factors constant. In Hong Kong's trading industry, the estimated model suggests that, based on the historical data, the wage rate is much controlling the labor demand, followed by the unit cost of capital. Meanwhile, the labor can be a substitute for capital in some extent as the estimated labor demand is increased by about 0.29% on average when the unit cost of capital is increased by 1%, holding other factors constant. For the output level (Y), in view of its low t ratio and coefficient value, it seems that its effect on labor demand is economically and statistically insignificant. For the major events, two events are statistically significant to bring the negative impact on labor demand as having the negative value in coefficients and significant t ratios. Worst still, the outbreak of global financial tsunami in 2008 brought the largest negative impact on the trading industry. Based on the estimated model's result, the aftermath decreased the estimated labor demand by about 0.1074*100%=10.74%, holding other factors constant. The result is incredible but reasonable as the global financial tsunami in 2008 directly affected the individual firm to decrease their output level due to the significant drop in demand and the economic recession. When the output level is lower, the labor force needed is dramatically decreased at the same time. This quantitative result provides significant evidence that Hong Kong Monthly Digest of Statistics had mentioned in the May 2011 issue, which discussed in Section 4.

6. Conclusion



Figure 2

In regard to the prediction performance, from Figure 2, the estimated model (10) can capture the similar pattern as the actual growth in labor demand from 1981 to 2009. More importantly, by proposing the dummy variables to model the major events appeared in 2001, 2002 and 2008, the prediction performance is significant and the estimated model almost captures the same pattern during these periods. It is evident that the dummy variables are useful to improve prediction. The graphical result in Figure 2 has validated the high goodness of fit shown in the high adjusted R^2 , which discussed in Section 5. However, when using the estimated model to predict the labor demand in 2010, it seems that the prediction performance is not quite significant than before as the predicted value cannot capture the same growth as the actual value. The reason behind this poor prediction performance is probably because of the aftermath of the global financial tsunami in 2008. The trading industry in Hong Kong was still recovering from the economic recession and thus the actual growth in labor demand in 2010 is close to zero. Going further to 2011, the estimated model seems to capture the same increasing growth rate as the actual value because the economic situation becomes stable and the industry has already recovered from the recession. This implicates the sudden event and its aftermath will adversely affect the prediction accuracy and increase the modeling difficulty. Unfortunately, this is inevitable in the real world and difficult to control and avoid unless assuming the sudden event that has the similar effect on the previous events in 2001, 2002 and 2008 and making use of the dummy variables in the estimated model (10) to improve prediction.

To conclude, the estimated model (10) can provide more insight into the relationship, especially the quantitative relationship, between labor demand, several factors and major events in Hong Kong's trading industry. This enhanced understanding will be conducive to the labor economy and different parties, such as the government, politicians, economists and firms in Hong Kong.

Last but not least, for any interested researchers, further improvement can be made in (5) discussed in Section 3. In fact, several assumptions have been made to formulate the econometric model (5). For the simplicity and modeling purpose, (i) the whole industry is assumed to be an individual firm in Hong Kong; (ii) the factor inputs and output are assumed to homogenous; (iii) two input factors are assumed to produce the output in the trading industry; (iv) Cobb-Douglas production function is used to describe the production process in the trading industry; and (v) the conditional demand function is used instead of the unconditional demand function due to the availability of data. To make further improvement, several points can be considered: (a) Constant Elasticity of Substitution (CES) production function can be used to formulate the conditional labor demand function in Section 2; (b) multi-input factors can be considered to produce the output; and (c) from Engle and Granger (1987), the Error Correction Model (ECM) may be applied to represent the short-term and long-term relationship if all I(1) time series data are cointegrated. Moreover, the in-depth analysis in labor demand in long run has been done by Hamermesh (1986), in which several modeling methods have been applied and discussed. It also has provided comprehensive investigation in labor demand in different industries.

References

Breusch, T. S. (1978)"Testing for Autocorrelation in Dynamic Linear Models," Australian Economic Papers, 17, 334-355.

Dickey, D. A. and Fuller, W. A. (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of American Statistical Association*, 74, 427-431.

Dickey, D. A. and Fuller, W. A. (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057-1072.

Durand, D. (1937)"Some Thoughts on Marginal Productivity, with Special Reference to Professor Douglas' Analysis,"*Journal of Political Economy*, 45, 740-758.

Engle, R. F. and Granger, C. W. J. (1987) "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.

Fuller, W. A. (1976) Introduction to Statistical Time Series, Wiley, New York.

Granger, C. W. J. and Newbold, P. (1974)"Spurious Regressions in Econometrics," Journal of Econometrics, 2, 111-120.

Hamermesh, D. S. (1986)"The Demand for Labor in the Long Run,"Handbook of Labor Economics, 1, 429-471.

Varian H. R. (1978) Microeconomic Analysis, W. W. Norton & Company, New York.