This is to certify that the

dissertation entitled

Optimal Port Congestion Charges and
Investment In Port Klang

presented by

Nor Ghani Md. Nor

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Economics

Kenneth D. Boyer
Major professor

Date Oct. 29, 1996

OPTIMAL PORT CONGESTION CHARGES AND INVESTMENT IN PORT KLANG

By

Nor Ghani,Md. Nor

A DISSERTATION

Submitted to
Michigan State University
in partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1996

ABSTRACT

OPTIMAL PORT CONGESTION CHARGES AND INVESTMENT
IN PORT KLANG

By

Nor Ghani,Md. Nor

This research highlights the need for considering changes in berth performance, as occupancy rate varies, in determining optimal congestion charges and investment in port facilities. A fairly significant empirical evidence, for different subsystems of berths, is presented to show that berth performance declines as the port gets more congested. Other ship and cargo characteristics were also shown to have significant impact on waiting time. In addition, service time variance was also studied in order to see if it is related to factors that have impact on the magnitude of service time. Optimal congestion charges and investment patterns were then calculated for Port Klang, the biggest port in Malaysia.

A theoretical model is developed to modify the existing one in order to enable us to incorporate changes in berth performance as port occupancy rises. This model is an extension to the current queuing model as it is applied to congestion charges and investment pattern analysis. The need for this model arises because, as it is now, the queuing model when applied to congestion and investment analysis fails to recognize changes in berth performance as the rate of berths occupancy changes.

ACKNOWLEDGEMENTS

I would like to express sincere appreciation to the members of my guidance committee, Prof. Kenneth Boyer, Prof. John Strauss, Prof. Bruce Allen and Prof. Lloyd Rinehart for their helpful advice and counsel during the course of completing this research. I am especially indebted to Prof. Kenneth Boyer for his indispensable role as advisor.

I would like to acknowledge the assistance of the management of Klang Port Authority, Klang Port Management Berhad and Klang Container Terminal Berhad for allowing me access to their database and a comprehensive tour of their facilities. I want to mention Mr Abdul Rashid Shukor, in particular, for his unparalleled interest and support during my one month stay at Port Klang.

My four year stay at Michigan State University was entirely funded by the Universiti Kebangsaan Malaysia, to which I am grateful.

Last, but not least, I would like to thank my wife and my three children, who have sacrificed a great deal to see me through this challenging process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This study is about how to improve port efficiency. Obviously there are several sources of *economic* inefficiencies at sea ports. One potential source is incorrect port charges. Yet another source of inefficiency is an investment decision which results in too much or too little investment relative to what is optimal. The wrong choice of technology mix is another possibility, but this generally falls within the realms of technical inefficiency.

Theoretically, the way to achieve efficient pricing is to set prices to equal marginal costs including the congestion costs, and the efficient level of investment equates the marginal cost of an additional unit of investment to the marginal discounted future streams of benefits coming from that investment. Mohring (1994) proves this point elegantly in the case of transport facilities like the sea port. However, there are many practical reasons why the theoretically correct pricing or investment levels in the ports are not achievable in practice (Walters and Bennathan, 1979). Some of these reasons include:

a) The concept of marginal cost may be hard to measure in practice.

b) The future benefits of marginal investment are hard to predict with a high level of precision.

c) The specification of the model representing port operation (which is used to guide pricing and investment decisions) may be deficient.

d) The capital market may put constraints on the financing options available to the port authority such that the port may have to resort to non-optimal pricing decisions.

This dissertation will look into some of the problems mentioned above using Port Klang, the biggest port in Malaysia, as the place of study. Unless Port Klang is hopelessly unique, the lessons derived from this research should be transferable to the study of port pricing and investment in general.

This research proposes a modified analytical model to overcome some of the weaknesses of the standard queuing models as they are applied to port. Getting a good estimate of the cost of congestion in port is very important because it directly affects optimal pricing and investments. To the extent that a standard queuing model (such as in Wanhill (1974) and Plumlee (1966)) has been used to guide congestion pricing and investment decisions, any deficiency in the model may result in non-optimal policy recommendations. Previous literature has indeed been hampered by arbitrary assumptions made on the standard model that lead to congestion variables like queuing time being calculated on unrealistic conditions. Our study will show that there is a deficiency in the specification of the standard queuing model and a more realistic specification will be proposed. In particular, the standard assumption that the service rate is independent of the port occupancy rate cannot be supported by the data from port Klang. Rather, the service time decreases as there are fewer ships in port simultaneously. In an n-berths system, for example, it will take longer on average to load and unload a ship when n_i berths are occupied then when only n_j berths are occupied simultaneously (where $n_i > n_j$). By showing this, we can conclude that the standard model leads to congestion charges that are lower than the level that is optimal. The implication on efficiency in both static and dynamic sense is clear. Since the existing literature has gotten the price wrong, the inefficiency can arise not only from the violation of the marginal cost pricing rule, but also from the error in the investment rules. For example, if demand is growing, the standard model guides us to investments in port that are "too little too late" because the congestion cost is always understated.

In order to find the optimal congestion toll and investment pattern, we develop an analytical queuing model that combine two characteristics. The model has to be a multi-channel or multi-server queuing model with a state-dependent service time. We need the multi-server feature because Port Klang is a multi-berth facility, and we require the state-dependent characteristics in order to incorporate the hypothesis that the port average service time increases as the port becomes more congested. State-dependent here means that the model has variable average service time dependent upon the number of berths that are occupied simultaneously. This combination makes the model harder to develop and more complicated, but this is needed so that we can make a better prediction on the congestion costs. Using the model, we derive explicitly several important variables on queuing time and costs so that we can find the optimal congestion toll function and the investment patterns.

To our knowledge there has been no paper written on port (or for that matter any service) facilities combining these two features. We have papers written on a single server facility with state-dependent service time (Lippman and Stidham, 1977) and multi-servers facility with fixed average service time (Knudsen, 1972 and Yechiali, 1972). So, our analysis will be a further extension of the existing literature. We do not come up with an entirely new way of analyzing the issue but our contribution will be in the form of modification and extension that place the two features in the same model and demonstrate what they mean to the issue of congestion pricing and investment.

Another goal of this study is to quantify the impact of the numerous features of ships and cargoes on ship waiting time. We postulate that different sizes and characteristics of ships and cargoes will result in systematic differences in the average loading and unloading time causing a different marginal impact on the queuing time and cost. This part of the study is important because when it comes to the actual determination of the toll schedule, charges must reflect the contribution of a particular ship to the waiting time. If the amount of cargo has a significant impact on the average service time, then we will need to quantify the impact of longer average service time on the average queuing time in order to arrive at the correct congestion toll for the different ships. We will certainly try to quantify the impact of cargo and ship characteristics on service time in the empirical section of the research. However, we only incorporate dependence of service time on port occupancy level in the theoretical model. Incorporating the other ship and cargo characteristics into the queuing model is undoubtedly a very important area for further research.

We also adopt a dynamic approach to the queuing system. Dynamic means that we allow ships' operators to balk when the cost of loading/unloading at the current port is 'too' large relative to an alternative port. It is important that we allow this response to ensure that the benefit of congestion tolls be realized. If ships' operators are not allowed to change their behavior, the congestion tolls will only involve a transfer of surplus from the ship operators to the port authority, and a toll does not have the intended discouraging effect. We can compare this approach to the one taken in the highway bottleneck congestion literature as in Arnott et al. (1990) and as in the highway flow congestion literature like in Ben-Akiva et al. (1986). The response of the commuters, however, were in terms of rescheduling the time they start leaving homes. This kind of rescheduling is not feasible in our model because we have a completely random arrivals, and more importantly, there is no preferred arrival times for all or some of the ships.

The final goal of this research involves empirical estimation of the optimal toll charges and timing of investments. The state-dependent multi-servers model will be used as a means of estimation. In order to perform the analysis, we need data on some key variables like the average service time for the different levels of occupancy or congestion index ‘c’, the arrival rate, the waiting cost, and the cost of constructing new berths. We will also require some projections of the future demand for port services.

This study is important because inefficiency in port can be very costly. The significance of sea ports in economic activities can be appreciated by looking at the sheer quantity and value of goods that pass through the sea ports around the world. In 1993, the annual value of goods traded internationally between nations amounted to $3.63 trillion (World Economic and Social Survey, UN, 1994); and a proportion of these goods are carried through sea transportation. Since ports act as the interface between trading nations and between sea and land transportation, inefficiency in the port operation can be costly given the volume of trade and the resources devoted to it.

Problem Statement and Hypothesis

The *standard* queuing model is one method that has been used by economists to represent port operations. The standard model is attractive because of its relative simplicity. In the area of pricing and investments, some research has focused on the general question of the reliability of this model as a vehicle to arrive at optimal congestion tolls and port investment. The reliability of the model has been verified by empirically testing its assumptions. A Poisson process for ship arrivals and exponentially distributed service rates (i.e. the probability distribution of time taken to load or unload a ship) are the two fundamental assumptions of the *standard* model which have been tested by several researchers. For example, Jannson (1982), Plumlee (1966) and Nicolaou (1967) conducted tests whether ship's arrival process is really Poisson. Jannson (1982) tested the assumption that the service rate is exponentially distributed and whether the service rate is constant regardless of the level of port utilization. All of the studies above generally confirmed that the arrivals of ships can be described by the Poisson probability distribution and the service time can be satisfactorily represented by the negative exponential distribution.

However, one unsatisfactory feature of the standard model is the assumption of independence between the service time and port occupancy rate. This assumption implies the average time taken to service a ship does not depend on whether the berths are relatively empty or highly occupied. It turns out that this assumption is very fundamental and relaxing it will affect the model in a very significant way. While this assumption of independence between the service and occupancy rates may sound reasonable when modelling queuing systems with departmentalized servers (like the telephone switchboards), we do have compelling arguments that the berths in port are not departmentalized servers. One of the goals of this research, then, is to prove that the standard assumption cannot be supported by data from Port Klang.

There are at least three reasons why we would expect the average service time of ships to be dependent on occupancy rate.

Firstly, it is unlikely that port activities are departmentalized. A non-departmentalized facility is one where the activity on one berth has influence on the performance of the other berths. If there is a more or less fixed pool of stevedores and cranes which can be moved between berths, then the more berths are occupied, the smaller will be the input of cargo handling labor and equipment per ship. So, we expect the service time per ship to rise as occupancy rate rises.

Secondly, handling of cargo in the storage facilities may also be affected when demand is high. As the storage facilities become more crammed, movements will become slower and more laborious and the average distance of internal transport will become longer resulting in reduction of throughput capacity of the port. This will cause the average service time of ships to increase as the occupancy rate rises.

Thirdly, as Walters and Bennathan (1979) argues, when there are more ships in port the waterways will become more crowded. Since in some ports lighters are an important part of the loading and unloading process, the fact that there are more lighters moving about means that manoeuvrability in port will become more difficult and slower when more ships are present in port. This will increase the average total service time required for ships to load or unload.

Another unsatisfactory feature of the queuing model as it is applied to port study today is the neglect of the potential impact of ship or cargo characteristics on the average service time. As a result, ships are not charged different congestion tolls, as in Rue and Rosenshine (1981), even though there may be systematic differences in average waiting time arising from different ship and cargo characteristics. It is fair to assume that a ship carrying, for example, a larger quantity of cargo will require a longer service time. Therefore, it should pay a higher congestion toll because by staying longer at berth, it imposes a greater externality to subsequent arrivals. The challenge, then, is to quantify the relationship between ship/cargo characteristics and service time. If all ships are required to pay the same amount of toll, which is the implication if we do not take cargo size or characteristics into consideration, than we will over-charge ships with smaller cargo and under-charge the opposite.

Importance of research

It is true that the standard queuing theory offers relative simplicity that may outweigh the advantage of a more accurate model, especially during the time when computer simulations are either too expensive or even impossible to do. However, the results derived from the queuing model (e.g. queuing cost), are very sensitive to the assumption of independence between service and occupancy rates. Jansson (1982, pg. 40) remarks:

"...even a fairly weak relation between 's' (service rate) and 'u' (the rate of capacity utilization) would invalidate the results obtained by standard queuing models."

On page 50 in his book "Port Economics" he further states that:

"...it can be shown that even modest positive relation between 's' and 'u' will result in substantially longer queuing times."

We have in fact tried to make some calculations to quantify the impact of the dependence between the service and the occupancy rate. We found that in the case of a two-berth system, the average waiting time increases by 140% if we have the service time increases by 25% whenever both berths are occupied. (In the calculation we have ships arriving at the rate of 10 per unit time and the service rate declines from 20 to 15 per unit time.)

Despite the implication that the assumption of independence has on the validity of prediction of the queuing model, so far there has only been one empirical research done to test the relationship between service and occupancy rates. The study was done by Jansson (1982), who found evidence to support the independence assumption.

From the theoretical standpoint, Jansson's finding is rather surprising *for port operations* for reasons we put forward in the last section. This research therefore, will serve as another attempt to find additional evidence for or against the standard queuing theory. If we find evidence against the assumption, we have a strong indication that the standard model needs to be re-evaluated. On the other hand, failure to find evidence against the standard model would provide additional evidence to support it. In fact, we would have produced a *stronger* evidence because of the improvements that we hope to make on the research done by Jansson. These improvements will be described in the methodology section.

Along with the attempt to test the interaction between the service and occupancy rate, we identify other factors that may systematically affect the average service time of ships. One such factor is the heterogeneity of ship and cargo characteristics. The quantity of cargo loaded/unloaded is another obvious possibility. Since different expected service times has a varying marginal impact on the total queuing time, the heterogeneity of ship and cargo characteristics should be taken into consideration in calculating optimal tolls. Thus far, ship and cargo characteristics have been ignored in research involving queuing model. We intend to fill this gap in the literature.

Another contribution of this research comes from the multi-server state-dependent model that it develops. This model is needed so that we can move closer towards estimating the correct congestion toll and investment pattern. In addition, the model can be used for other type of service facilities that exhibit dependence between the service and servers utilization rates.

Finally from the policy perspective, this research has a direct implication on the actual congestion pricing and investment decisions. We show that the difference is not merely theoretically significant but also practically evident. This is especially true since, as we stated earlier, even a relatively mild relationship between ‘u’ and ‘s’ will have significant impact on the length of queue and queuing costs. If our hypothesis is correct, then we could show that the port congestion charges would have been too low. In addition, when demand is rising, investment will be “too little too late” if the standard model is used as a vehicle of analysis. The impact on welfare is fairly evident. It will be improved because we will have a more accurate measure of the expected queuing cost by using the modified model.

Significant prior research

Papers written on queuing theory as applied to ports have been few and far between. Some of the earlier application of the queuing theory on ports can be found in the engineering literature. They were concerned with the question of optimum port size and the test of whether ship arrivals conform to the Poisson process. Examples of papers in this line of research includes Nicolaou (1967), Plumlee (1968) and Wanhill (1974). Nicolaou used the chi-square test to confirm that ship arrivals in the ports of Limassol and Paphos (both in Cyprus) follows the Poisson distribution. He also presented a graphical analysis of how investments in port can be achieved based on the criteria that the optimal port investment is achieved when the marginal benefit of reduced port congestion is equal to the marginal cost of that investment. The Plumlee paper tested the Poisson assumption using data from several ports in South America. It was found (as in Nicolaou) that the ship arrival pattern can be represented by the Poisson distribution. Plumlee also showed how we can determine the correct level of investment in port. The difference between his approach and Nicolaou is that Plumlee minimized the combined cost of ships waiting time and the cost of vacant berths. The investment level that minimizes the total cost is the optimal capacity for the port. Weille and Ray (1974) used the standard queuing model to show that the optimal investment level is achieved by weighing the benefits of reduced waiting time as a result of additional investment against the cost of making that investment.

None of the papers mentioned above has analyzed the issue of optimal port size based on optimal dynamic queuing behavior. In these models ship operators are not allowed to balk and there are no congestion tolls imposed. Our model assumes that an optimal congestion schedule must expect and make allowance for balking by ship operators.

Several papers have dealt with the idea of using congestion toll in facilities that exhibit queuing behavior, such as a port, to achieve a social optimum. Rue and Rosenshine (1981) show how toll can be used as a viable instrument to achieve social optimum in the case of exponential queuing with single channel. Their result explicitly demonstrates that in order for congestion toll to work, balking must be allowed. This is similar to the result in the dynamic highway congestion model where arrival patterns are altered by the congestion toll. Yechiali (1971) considers a similar problem with one major modification where he uses a general arrival pattern instead of an exponential one. In addition, there are two economics papers delineating the application of queuing theory in order to arrive at optimal congestion toll by Naor (1969) and Edelson and Hilderbrand (1975). These two papers are general theoretical papers which are written for general service facilities that have queues as a feature. Both papers argue that facility charges must reflect the marginal congestion cost if we want efficient use of resources and they show theoretically how this can be done.

Several extensions and variations to Naor's model can also be found in the literature. Knudsen (1972), considers a multi-server queuing model but with a fixed waiting time as a result of additional investment against the cost of making that investment.

None of the papers mentioned above has analyzed the issue of optimal port size based on optimal dynamic queuing behavior. In these models ship operators are not allowed to balk and there are no congestion tolls imposed. Our model assumes that an optimal congestion schedule must expect and make allowance for balking by ship operators.

Several papers have dealt with the idea of using congestion toll in facilities that exhibit queuing behavior, such as a port, to achieve a social optimum. Rue and Rosenshine (1981) show how toll can be used as a viable instrument to achieve social optimum in the case of exponential queuing with single channel. Their result explicitly demonstrates that in order for congestion toll to work, balking must be allowed. This is similar to the result in the dynamic highway congestion model where arrival patterns are altered by the congestion toll. Yechiali (1971) considers a similar problem with one major modification where he uses a general arrival pattern instead of an exponential one. In addition, there are two economics papers delineating the application of queuing theory in order to arrive at optimal congestion toll by Naor (1969) and Edelson and Hilderbrand (1975). These two papers are general theoretical papers which are written for general service facilities that have queues as a feature. Both papers argue that facility charges must reflect the marginal congestion cost if we want efficient use of resources and they show theoretically how this can be done.

Several extensions and variations to Naor's model can also be found in the literature. Knudsen (1972), considers a multi-server queuing model but with a fixed average service time. Lippman and Stidham (1977) extended the model to include general service time instead of an exponential one. The reward to queuing customers are assumed to be random. Another paper by Yechiali (1972) uses a general arrival pattern instead of an exponential one but includes a multi-server, instead of single-server, feature. None of these papers have dealt with the case we are currently considering, namely the situation when we have multiple berths with variable performance depending on occupancy rate. Note also that congestion issues were considered only in the *theoretical* papers of the above mentioned researches.

The *only* empirical research that employs the standard queuing theory with a direct treatment of port congestion pricing is one by Jansson (1982). In his research, service time and occupancy rate are found to be independent. Our research will re-evaluate this finding using a more detailed data and analysis.

In his study, Jansson regressed occupancy rate against port throughput using linear and exponential specifications. If congestion exists in ports, increases in port throughput should correspond to a progressively rising port occupancy rate. He argued further that his study therefore provided an indirect test of the assumption of independence between service and port occupancy rates because as throughput rises occupancy rate should only rise proportionally if the assumption is correct. His data support to the standard theory even though (as we have argued earlier) there are reasons to believe otherwise. He tried to explain this result by saying that what might have happened is as the port got more congested shipping companies or the port authority hire more stevedores to work overtime. As a result there was no drop in average throughput as occupancy rate rises. However, his conjecture was not supported by the data. When he regressed port throughput against stevedoring costs he *did not* find the stevedoring cost increasing progressively as throughput rises.

Our conjecture is that there are two reasons why Jansson's research found evidence in support of the independent hypothesis between service and occupancy rates. First, his monthly averaging period is relatively long. Values at the high end of the spectrum are "averaged-out" through the averaging process even though these are the times when the port were most congested and expected to deliver a lower service rates. Schulze (1978) explains this averaging effect. To avoid this problem, in our study, we will instead use individual ship data.

Second, Jannson also failed to control for the types of ships that were in port. Large bulk and container ships are known to have a higher rate of being loaded and unloaded (see Robinson, 1978 for evidence at the port of Hong Kong). Since these ships are known to have higher loading and unloading capacity, port throughput was higher when there were more bulk and/or container ships in port, given occupancy rate. In addition, really high throughput is only achievable when there are many of these ships in port. As a result Jansson was not able to find higher congestion in port when throughput is high because larger throughput is explained, at least partially, by the existence of this type of ships in port. So, his test is not a proper test of independence between 's' and 'u' because of the built-in congestion-reducing effect of container and bulk ships. In other words, if ships were homogeneous, Jansson will find more congestion in port when throughput is higher. Our study will correct this problem first by grouping the berths according to the kind of ships they serve and second by having a more direct test using a direct measure of service and occupancy rate. To summarize, our method tries to improve upon the research done by Jansson in three different ways:

1. More direct measures of ship occupancy rates and service times are used.
2. Additional variables, like type of ship and type of cargo, to control for variations in service time which are not accounted for by the occupancy rate is introduced.
3. Individual ship data are used. This allows for fluctuations in the values of the variables to better approximate the real situation in port and avoid the smoothing effect of the averaging procedure.

Having described some of the important literature related to our research, we move next to outline the methodology of this research.

Chapter 2

Port Klang, Malaysia: The Institutional Settings

Introduction

The biggest port in Malaysia in terms of both throughput tonnage and number of ship calls, Port Klang is located about 60 miles from Kuala Lumpur, the capital city. It lies at a strategic location along the major shipping route serving ships sailing between the Far East and Europe. It is also the port for the Klang Valley industrial area in Malaysia.

The port started operation in 1890 with three wooden jetties (KPA, 1988). Several years later other wharves were built through which the British colonialists exported rubber and tin to England and other parts of the world. From its humble beginning of three wooden jetties, the port grew a present total of 25 modern berths serving different kinds of cargo. The berths can generally be divided into several sub-systems. There are container, breakbulk, liquid bulk and dry bulk sub-systems of berths, both for import and export cargo. In 1994, the port served more than 6000 ships and handled 33.8 million tonnes of cargo. The port is also expected to grow at the rate of 10% per year in terms of cargo tonnage and 6% per year in terms of ship calls in the next five years (Abdul Kadir, 1994).

This chapter describes the important institutional features of Port Klang that are likely to have some bearing to the pricing and investment decisions. Understanding the institutional features is very important because they can have significant impact on pricing and investment. The chapter is divided as follow:

A. The Port Authority and The Operators: Managing the port.

B. Privatization, Port Charges and Investment Policy.

Under heading (A) we describe the organizational structure and the functions of the port authority as well as its relation to the private operators. Part (B) gives an overview of the privatization program, the process of setting port charges and the way in which investment is decided at the port. We also offer some discussions on the likely impact of the various institutions and arrangements on pricing and investment decisions. Many of the points mentioned in this chapter were discovered during our visit to the port through interviews with the port officials.

A. The Port Authority and The Operators: Managing the port

The present organizational and ownership structure owes much to the Port Authorities Act, which was passed by the Malaysian parliament in July 1963 and the privatization effort beginning in 1986. Prior to 1963 the port was handled by the Malayan Railway (KPA, 1992 (i)). Since 1963, the port has been managed by the Klang Port Authority as a public port. The day to day running of the port is done by the management team headed by a general manager. The overall planning and policy making function is accomplished by the Board of the Port Klang Authority with members appointed by the Minister of Transport. Working directly under the Board are several committees that plan for the various aspects of the port operations. These committees are Finance and Trade Facilitation Committee, Planning and Development Committee, Service Committee, and the Port Consultative Committee. The Port Consultative Committee is really a council that serves as a place where port users and ancilliary service providers are able to voice their opinions and concerns and ultimately influence the decision of the port. This committee has 21 members serving diverse interests.

We mentioned earlier that the Ministry of Transport wields considerable influence on the port through appointing members to the Board of the KPA. However, it is not a passive entity with its job limited to appointing the Board of Directors of the port. It has the final authority to approve or disapprove changes in rates schedule and investments exceeding M$1 million (US$400 000). More importantly, it does at times make decisions that have a strategic impact on the organizational structure and operations at the port with political consideration weighing heavily in their decision making process. Two examples show that the port was affected significantly as a direct consequence of a political decision by the Ministry. The important point is that these decisions were made based on external considerations quite unrelated to the port operation. The first example comes from the early sixties. The Port Authorities Act (1963) was drafted following the repeated strikes conducted at that time by the railway union members. Many times the port had to be closed down because of railway workers union strikes arising from matters unrelated to the port. In order to reduce the effectiveness of the railway strike (because the strike caused the port to cease operation), the management of the railway and the port were separated in 1963 by the creation of the Klang Port Authority. A more recent example occurred when the Ministry of Transport decided to privatize the port as part of an overall government program towards privatization. We argue later that the choice of which berths were to be privatized first was also probably politically motivated. These examples serve to demonstrate that general political issues can and do have an impact on the operations and organizational structure at the port.

After privatization, Port Klang can be best described as a regulated port with private corporations running the daily operations. The KPA no longer functions as a port operator but is now a landlord-cum-regulator. Currently, the wharves, the fixed structures, and about 800 hectares of land are owned by the Port Authority, a federal body created by the 1963 statute (KPA (i), 1994). Through the federal government privatization program, these wharves and fixed structures have been gradually leased out for a contract period of twenty one years to private operators. Two corporations were awarded the lease contract. Klang Container Terminal Limited (KCT) manages four container terminals identified as wharves 8, 9, 10 and 11. The remaining wharves are leased to the Klang Port Management Limited (KPM). The management of the corporations are typical for any private corporations.

Privatization at the port has to be understood carefully since it does not mean an outright sale of the port to private corporations. Neither do the private operators have complete freedom over pricing and investment decisions. Rather it is a mixture of sale and leasing of assets with regulated pricing and investments. It may be interesting to see the impact of these arrangements on the pricing and investment behavior of the private operators and the port authority. These issues, are discussed in the privatization section of this chapter.

Constructing new wharves and buildings is still under the jurisdiction of the port authority. The lessees which actually operate the port can, of course, make the initial proposal to the port authority for investment in new wharves and buildings, but they are not authorized to construct their own buildings and wharves. They are only free to replace or purchase movable assets (like cranes and straddle carriers) subject to safety clearance from the authority. Other responsibilities of the port authority include ensuring security, providing fire-fighting and commissioning dredging work on the approach channels (KPA, 1992 (ii)).

The authority is also responsible to oversee that rates approved by the ministry are implemented by the lessees. They may also recommend to the ministry changes in charges for port services and investments in new wharves and fixed structures as it sees fit or upon request from the private operators. There are no periodic (e.g. annual) reviews of port charges. This by itself is interesting since one would expect that fluctuations in cost of providing port services would demand that the rates be reviewed on a regular basis. We were not given access to the lease documents to ascertain whether there are clauses that describe the procedure of rate changes. Our information is limited to the statements made by the Port Authority's, KCT's and KPM's officials, who indicated that the operators can ask for changes in port charges.

The private operators own the movable assets and are responsible for maintaining the wharves and providing port services to users. In exchange they are allowed to levy charges on port users based on the rate schedule approved by the port authority.

The day to day management of loading, unloading, transportation and storage operations are done by the port operators. They are also free to compete with each other even though competition is restricted to the non-price aspects of the port services. This is because the rate schedule is determined by the port authority with the approval of the Ministry of Transport. However, the degree of adherence to the rate schedule is unknown and it would be interesting to find out whether transaction prices match regulated prices. One can certainly envisage the mounting pressure for both operators to compete in price in a period of low demand. This could be done, for example, through hidden rebates during an economic recession.

Marketing efforts for the port are done by the port authority as well as by the individual port operators. The authority generally focuses on promoting the port as a whole. It is fairly common for the port authority to bring along representatives from KPM and KPA on their promotional trips. However, the individual operators tend to direct their effort towards attracting a particular shipper or shipping lines to call at their leased facilities.

There is a continuous communication and close co-operation between the port authority and the two port operators. For example, they plan and go on promotional trips together. Both operators also supply weekly reports of cargo and ship data to the port authority. On one occasion, the port authority also acted on behalf of the operators, in petitioning the Custom Department to speed up the custom clearance terminal. It had been alleged that the terminal was a chronic bottleneck for the flow of cargo in and out of Port Klang. It is not clear, however, whether the two operators collude in the container terminal business. It is certain that no shipping line uses the services provided by both operators, but that by itself does not tell us enough. (Recall that KCT only manages container terminals while, KPM handles the operation of different terminals including container terminals and that collusion to divide customers and markets is common in other contexts.)

Another interesting managerial and operational feature of the port is that the pilotage service is provided only by KPM. Pilotage refers to the navigating service that is used as an aid to the captain of a ship to ply the water in a defined area around Port Klang. Pilotage is compulsory for all ships entering and leaving Port Klang. Therefore, even those ships intending to berth at wharves belonging to KCT will have to engage the pilot service of KPM. One wonders how faithfully the first-come first-served rule is adhered to, especially during a busy period. However, the port authority confirmed that there has never been any complaint from the KCT management regarding the pilotage service.

B. Privatization, Rate Setting and Investment Policy

The privatization of Port Klang is really part of a bigger privatization program that the Malaysian government launched. Beginning from the mid 1980's, it involved many state-owned corporations. In fact, Port Klang is the first state-owned entity to be privatized. Remember that privatization was not an outright sale of the port to private corporations. It involved only the operational side of the port activities while the land, wharves and the fixed structures were leased for a period of twenty one years. The movable assets on the wharves, like the cranes and the straddle carriers, were sold to the corporations that were awarded the contract. When asked about the reason for privatization, port officials often cited the goal of improving operational efficiency as the main motive. Cumbersome bureaucratic process were required before privatization. The requirement that purchases exceeding M$1 million (US$400,000) must get a prior approval from the Ministry of Transport was the favorite example. This meant that the ministry must approve even the purchase of a crane, which runs into several million Malaysian dollars. In addition compensation scheme was also not flexible then, because employees of the port were civil servants working under a rigid pay structure, with a secured employment status that affected work habits and productivity.

The container terminals were the first group of wharves to be privatized, in 1986 (KPA, 1988). Nothing in the publicly available port literature indicates why the container terminals were privatized first. However, several interviews with the port official suggest that privatization occurred here first because those terminals were very profitable. Anxious to ensure success to its overall privatization program, the government, arguably, needed a good start. This made the container terminals a natural choice. It is also of interest to note that when the container terminals were privatized, the Port Authority held a 49% stake in the KCT, the private company that was awarded the contract. Over the years the stake has been reduced, and it now stands at 20% (KPA, 1991 and KPA(ii), 1994). This was not the case, however, for the privatization of the other wharves. When the remaining wharves of the port were privatized, in 1992, the Port Authority held no stake in the company that were awarded the contract i.e. KPM. Again we can find no official explanation for this fact from the publicly available port literature.

The nature of the lease agreement also deserves some discussion. In the case of KCT, the lease requires that an annual fixed sum is paid to the port authority with a predetermined increase of 5% every three years (KCT, 1994). Although the amount of lease payment is different, the same formula is applied to the lease contract made with the KPM. Apparently the lease is constructed in such a way that the operators will take all the business risks as well as reaping all its rewards. At the same time, the contract period is relatively long and we can suggest several reasons for it. First, the port authority may want to signal to potential lessees that it will not engage in opportunistic behavior because of the specificity of investment in port. Second, the authority may also want to encourage the lessees to look at the port business with a long term perspective and to engage actively in promoting the port because of the specificity of promotion expenditures and the long term nature of its impact. In the language of industrial organization theory, a long term contract (i.e. the twenty one year lease) is used to avoid under-investment arising from the specificity of assets and advertising expenditures (see Tirole 1988, for explanation of asset specificity).

Upon privatization, the employees of the Port Authority who were directly working on the leased or 'privatized' wharves were given the option of joining the new corporation. Except for the management team, the employees at the privatized terminals were generally the same before and after privatization because almost all of the former port employees opted to work for the new corporation. Compensation, work schedule, training and work habits may have changed after the privatization, however.

Another important feature of this privatization effort is that the operators are not free to set their own schedule of charges. Charges are still technically set by the port authority after getting the approval from the Ministry of Transport, even though the operators can petition for rate changes. Rates are not set based on a formula (like a desired rate of return on capital employed) as would be expected for a regulated port. In fact rates that are in force today are really a historical construct. Since 1963, the rates have never been changed except for some minor amendments in 1977 and 1992. The rates formulated in 1963 are generally still intact. Surprisingly, the port authority confirmed that there has never been a request from the port operators for rate changes except for the minor amendment for charges to passenger ships in 1992. Procedurally speaking, the process can be put into motion when the port operators request to the port authority for a change in the schedule of charges. After reviewing the application and feeling satisfied that the request is justified, the port authority will then make a formal application to the Ministry of Transport for the final approval. It is not known, however, if the operators do really charge the published rates on the port users.

One may argue that the rate changes are actually made periodically because the parties to the contract will renegotiate the term of the lease. The counter argument however, is that twenty one years is really a long interval for rate changes to take place. Yet another plausible argument is that the actual charges are not really the published rates, but operators use different rates that suit the market condition. If that is the case then they are likely charging below the published rates. Otherwise, the port authority would have received complaints of overcharging because the schedule of charges are available to port users. Since no complaint has been received, it is very likely that the operators are not charging more than the published rates.

Is it possible that the operators prefer to have port charges regulated by the port? Price fixing is easier if it can be enforced legally through the leasing contract. However, there is no anti-trust law in Malaysia that will prevent them from colluding anyway. Furthermore, there are only two firms coming together, which means they can collude fairly easily. Presumably, they can perhaps find a trigger strategy, as can be found in oligopoly collusion theory, to ensure discipline between them to maintain a certain level of prices. However, regulation has one advantage over tacit collusion. It is certainly a more convenient tool for the operators to enforce discipline since any digression from the published rate will be illegal because it is a breach of the lease contract.

On the other hand, one can still argue that the reason for regulated prices is that the port authority really wants to use it as an instrument to achieve social welfare maximization. However, it is difficult to reconcile this assumption with the fact that port charges have generally been unchanged since 1963.

Investment is another important issue. Investment in fixed structures like wharves and buildings is under the jurisdiction of the port authority even after privatization. Investment in other types of capital, however, is generally carried out by the operators. The operators can, of course, make a request for investment in fixed structures; but even if the port authority decides to grant the request, the construction project will still be under its responsibility. Once the structure is completed it will be leased to the operator.

Now, will the investment pattern be different if the port operators instead of the port authority be ultimately responsible for investment in port? Well, firstly it depends on how divergent are the goals between the port authority and the operators. Assuming the port authority pursues social welfare maximization and the operators try to maximize profit, we may have different investment plan over time between the authority and the operators. Certainly, with leasing, the goal of welfare maximization on the part of the port authority will have to work within the constraint that the operators receive a reasonable return on their leases. Another reason why the investment pattern may be different if decision is left to the operators is that, it is more likely for fund borrowed by the port authority to carry a lower interest rate because of the lower default risk. As a result investments will be evaluated more favorably by the port authority given that the port and the operators both have the same kind of information on the profitability of the project. Yet another reason for probable differences in investment decisions, is that the port authority will be more likely to consider social costs in evaluating a project unlike the private operators. Finally, political consideration can also play a major role in investment decisions by the port authority in certain circumstances.

Budget constraint is another issue that can have a significant impact on the investment decisions at Port Klang. To the extent that the port receive funding from the federal government, means that decisions whether to invest in one project may not be based on the merit of the project itself. Other social priorities may supersede the requirement of the port capital funding. Government revenues and expenditures generally come through a common pool. This alone may result in certain worthy projects be rejected because of budget constraint. This constraint, however, is relaxed if the port authority can raise funding for capital project from the open market and leases back the asset to private operators to repay and service the debt.

Conclusion

This Chapter described the institutional features of the port and the likely implications of those features on the port operation in general. We put a lot of emphasis on the issues of pricing and investment within the established institutional settings, the role of the port authority, the private operators and the relationship between them. We also outlined the privatization program at the port and what it means to the operation at the port. Finally, we discussed the process of rate scheduling and capital investing.   The institutional features of the port partly explain the process and the major forces bearing on issues of pricing and investment. Decisions on pricing and investment are made within an institutional framework and not in a vacuum.

CHAPTER 3

State-dependent Multi-server Queuing Model

The Model

This chapter describes the state-dependent multi-server queuing model to be used to obtain the optimal toll and investment pattern in Port Klang. By state-dependent we mean that the performance of the berths in a sub-system is assumed to be a function of the degree of simultaneous berth utilization. For example, in a three-berth sub-system, state dependence holds if the average loading/unloading times depends on whether one, two or all of the berths are simultaneously occupied.

We introduce the model from the basic principles of queuing models in order to see how the assumption of dependence between port utilization and average service time is represented. There are several texts that provide excellent exposure to the mechanics of queuing theory. Among them are Taylor and Carlin (1994), Ross (1993), Allen (1990), Gross and Harris (1985), Cooper (1981), Cooper et all (1977), Newell (1982), Kleinrock (1975) and Gorney (1981). We especially benefit from the first few chapters (1, 2, 3 & 4) in Gross and Harris (1985). We also learned several algebraic 'tricks' that help in simplifying long and complicated expressions from this book. This 'tricks' are especially useful since queuing models are normally replete with long and involved algebraic expressions.

Much of the development for the beginning part of this model will be relegated to Appendix 1 because it involves a lot of algebraic manipulations. In addition, at the initial stage we develop the model without reference to the congestion issue. The subsequent part proceeds to see how the model can be used to derive some expressions useful in estimating congestion costs in port. The model is an extension of the model presented in Naor (1969). After its publication, this paper has been extended and reworked to incorporate different queuing systems with varying features. Nevertheless, despite these variations as in Lippman and Stidham, (1977), Knudsen (1972), Yechiali (1971 and 1972), Edelson and Hilderand (1974) and Rue and Rosenshine (1981), no one has tackled the situation with multiple servers and variable service times. We fill this gap while maintaining a similar technique in solving the optimal congestion toll. We also add a section to derive the optimal investment pattern based on the queuing behavior after the imposition of an optimal toll schedule. This additional section deals with an issue not previously considered in any of the papers listed above.

We begin with the assumption of an m-berth port system. We further assume in the beginning that the facilities at those berths are able to load/unload a ship at the rate of μ_1, μ_2 , μ_3.........μ_n when there are 1, 2, 3.........n ships in the system respectively. As in the standard queuing theory, we assume that the service rate per berth follows the negative exponential distribution with parameter μ . μ is defined as the expected number of ships serviced per unit time. The arrival rate per unit time, on the other hand, is assumed to follow the Poisson distribution with parameter λ. Our initial goal is to derive the probability density function for the different possible states in the system. States here refers to the number of ships in the system at any particular moment in time whether in queue or in service.

With the above assumptions we can show (see Appendix 1, Section A) that the state probabilities reduce to:

$$p_n = \frac{(\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0)p_0}{\mu_n\mu_{n-1}\cdots\mu_1} \qquad \textit{for } n \geq 1 \qquad (1)$$

Suppose further that we have 'c' servers or berths in operation and that the average service time is such that:

$$\mu_n = n\mu_1 \qquad \textit{for } c > n \geq 1$$

$$\mu_n = c\mu \qquad \textit{for } c \leq n$$

There is an important point to be made at this juncture. Berths' performance switches to a single different rate only when all berths are occupied. In this construction, ships are served at a lower rate per unit time when all of the berths are occupied, thus $\mu < \mu_1$. We could have assumed that the performance of the berths decreases gradually step by step, having more than two levels of μ i.e., different values for μ when there are one, two, three or more berths occupied. However, we greatly simplify the model by assuming two rates because the algebra gets geometrically more complicated when the number of applicable service times increases. Furthermore, several of the subsystems to be considered in the empirical section of this research only involve two berths, which makes the assumption appropriate. One rate applies when there is only one ship present and the other when there are two ships present simultaneously. In addition this assumption will give us the most conservative assessment on the magnitude of the impact of a berth's utilization on queuing time since we ignore the decline in berths' performance before maximum utilization is reached. This is useful since we are trying to show that ignoring changes in service time results in a significant welfare loss.

With these additional assumptions on the number of servers and declining performance when all servers are fully occupied, we can show (see Appendix 1, Section B) that

$$p_n = \frac{\lambda^n p_0}{n!\,\mu_1^n} \qquad \text{for } c > n \geq 0 \qquad (2)$$

and

$$p_n = \frac{\lambda^n p_0}{\left[(c-1)!\,\mu_1^{\,c-1} c^{\,n-c+1} \mu^{\,n-c+1}\right]} \qquad \text{for } c \leq n \qquad (3)$$

where

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{r_1^{\,n}}{n!} + \frac{r_1^{\,c-1}}{(c-1)!(1-\rho)}\right]^{-1}$$

and c is the number of berths in a sub-system.

μ is the service rate per ship per unit time when all berths are occupied.

λ is the arrival rate per unit time

$$\rho = \frac{\lambda}{c\mu}, \quad r_1 = \frac{\lambda}{\mu_1}$$

We can continue to find the state probabilities, p_n, by substituting p_0 into equation 2 and 3. Having calculated p_0 and p_n, we can then solve for the expected queue length and expected waiting time which are crucial in determining the congestion costs by employing the formula $L_q = \sum_{n=c}^{k} (n-c)p_n$ for the expected queue length. Subsequently, Little's formula[1] can be used to get the expected waiting time in the queue. Using this

[1] Little's formula is expressed as $L_q = \lambda W_q$. This formula can be explained rather intuitively. Consider a customer that has just joined the queue. On average he will have to wait for W_q before he steps into

information on the expected waiting time, we proceed to calculate the marginal waiting cost by first multiplying the ships' queuing cost per unit time with the total expected waiting time, and then differentiating the function with respect to n, the number of ships. By equating the marginal cost of waiting per unit time for an additional arrival, we arrive at the optimal congestion toll that we can impose on all ships arriving at the port. However, this approach will only provide us with a static analysis of the congestion toll since we neither allow ship operators to respond to the congestion toll nor allow them to balk when they arrive and find the expected waiting time 'too' costly. A more meaningful and realistic way to look into this issue is to have a dynamic analysis where ship operators are allowed to react to the imposition of toll as well as to varying expected waiting cost. If ship operators are not allowed to change their behavior, the congestion tolls will only involve a transfer of surplus from the ship operators to the port authority; and only when ships are allowed to balk do we have the intended discouraging effect of a toll. We can compare this approach to the one taken in the highway bottleneck congestion literature as in Arnott et all (1990) and as in the highway flow congestion literature like in Ben-Akiva et all (1986). The response of the commuters, however, was in terms of rescheduling the time they start leaving homes. This kind of rescheduling is not feasible in our model because we have completely random arrivals, and more importantly, there are no preferred arrival times for all or some of the ships. In our model reaction is allowed through balking.

The introduction of balking requires incorporating the decision making process and behavior of ship operators. This is precisely what the next section does.

---

service. Once he enters service the, average length of queue behind him now is L_q. It must have taken

Private Optimization Without Congestion Toll

We begin this section by describing the strategy of private ship operators when no toll is imposed. In deciding whether to join the queue, ship operators will weigh between the cost of loading/unloading at the current and the alternative ports. As long as the expected the expected cost is lower at the current port, the optimal decision, from the operator point of view, will be to join the queue. We further assume that if an operator decides not to queue and go to an alternative port, he will then incur an expected cost that is greater than the cost of loading/unloading at Port Klang when he arrives and immediately finds a berth for loading/unloading. Otherwise, he wouldn't have called to the current port in the first place. We call the difference between the expected cost of loading/unloading at an alternative port and the cost of immediate loading/unloading at the current port (excluding waiting cost), the penalty cost (F). The idea is that a ship operator will decide to join the queue at the current port as long as the expected cost of waiting is less than the penalty of unloading his cargo at an alternative port. In the case of Port Klang the alternative port is Port Singapore. We further assume that the penalty is common to all ships. This may appear unrealistic, but it greatly simplifies the analysis. Finally we assume that all ships incur the same cost of waiting per unit time.

Given our assumptions above, the strategy that is going to be adopted by a ship operator is to first observe the number of ships already in port either waiting in service or queuing for a berth. This number is a random variable. If it is less than the selected strategy, say k^p, then he will join the queue. Otherwise, he will go to an alternative port

---

$1/\lambda$ for each of the L_q to arrive, so that, $L_q (1/\lambda) = W_q$. This gives us the above expression.

and incur a penalty. The private operators' strategy, k^p, must satisfy the inequalities below:

$$F - k^p C_q\left(\frac{1}{c\mu}\right) \geq 0$$

$$F - \left(k^p + 1\right)C_q\left(\frac{1}{c\mu}\right) < 0$$

where: $c\mu$ is the number of berths multiplied by the service rate per ship per unit time.

k^p is the chosen strategy.

F is the penalty for berthing at an alternative port.

It is optimal for a ship operator to choose k^p such that the two inequalities above hold. The first inequality requires that when k^p is chosen the penalty (F) must be either equal to or higher than the expected waiting cost. The second inequality is the condition that, at k^p+1, the expected cost of waiting is higher than the expected penalty. These two inequalities can be rewritten so that we get $k^p \leq \frac{Fc\mu}{C_q} = Z_p < k^p + 1$. This condition essentially says that k^p is the largest integer not exceeding $\frac{Fc\mu}{C_q}$. Each ship will always join the queue as long as the system size, when it arrives in port, does not exceed k^p-1.

Another important point to note is that since all ships have the same cost of queuing per unit time, they will all pursue the same strategy, say k^p. Given this strategy we really need a queuing model with finite waiting space for calculating the socially optimal queue length. The finite waiting space feature arises not because of any physical constraints but from the strategy that is pursued by ships' operators. This is the approach taken by Naor (1969) and Rue and Rosenshine (1981) and Yechiali (1971) and we are follow a similar route. This feature is introduced in the next section which considers the strategy for social optimization.

Social Optimization

We now introduce the finite waiting space feature into the queuing model. A queuing model has finite waiting space if there is a limit placed on the size of an allowable queue. In this model, the arrival rate is specified as follows:

$\lambda_n = \lambda$ for $0 > n \geq k$

and $\lambda_n = 0$ for $k \leq n$, where k is the limit placed on the queue size.

By incorporating a limit on the allowable queue size, we get the following expression for the state probabilities, p_n.

$$p_n = \frac{(\lambda^n)p_0}{n\mu_1(n-1)\mu_1(n-2)\mu_1(n-3)\mu_1\ldots\mu_1}$$

$$= \frac{(\lambda^n)p_0}{n!\,\mu_1^n} \qquad \text{for } c > n \geq 0$$

and

$$p_n = \frac{(\lambda^n)p_0}{\mu_n\mu_{n-1}\cdots\mu_1}$$

$$= \frac{(\lambda^n)p_0}{[(c-1)\mu_1\ldots\mu_1][c\mu]^{n-c+1}}$$

$$= \frac{(\lambda^n)p_0}{(c-1)!\,\mu_1^{c-1}c^{n-c+1}\mu^{n-c+1}} \qquad \text{for } c < n \leq k$$

The two expressions above are derived from equation (1) from the earlier section. The two equations are essentially the same as before except that the range of n is limited to k instead of infinity. We also have a modified expression for p_0 and it can be shown (see Appendix 1, section C), that it is as follows:

$$p_0 = \left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!} + \frac{\rho r_1^{c-1}}{(c-1)!}\frac{(1-\rho^{k-c+1})}{1-\rho}\right]^{-1}$$

We also want to note at this point that the condition $\rho < 1$ for steady state is no longer required since the finite waiting space ensures that arrivals will be turned away once the capacity is reached. This avoids the problem that the expected queue length approaches ∞ as ρ approaches one. After finding p_n we are now ready to find the expected number of ships in the *system* (i.e. the average number of ships queuing as well as waiting for service at berths) at any given point in time. The difference between system and queuing size is the inclusion of the expected number of ships waiting at berth while being serviced. Following Gross and Harris (1985), it is generally easier to first calculate the queue length, L_q, before finding the expected system size. L_q is given by:

$$L_q = \sum_{n=c}^{k}(n-c)p_n$$

$$= \sum_{n=c}^{k}\frac{(n-c)\lambda^n p_0}{(c-1)!\mu_1^{c-1}c^{n-c+1}\mu^{n-c+1}}$$

$$= \frac{p_0 r_1^{c-1}\rho^2}{(c-1)!}\sum_{n=c}^{k}(n-c)\rho^{n-c-1}$$

$$= \frac{p_0 r_1^{c-1}\rho^2}{(c-1)!}\sum_{m=1}^{k-c}m\rho^{m-1} \qquad \text{where } m = n\text{-}c$$

$$= \frac{p_0 r_1^{c-1}\rho^2}{(c-1)!}\sum_{m=1}^{k-c}\frac{d\rho^m}{d\rho}$$

$$= \frac{p_0 r_1^{c-1} \rho^2}{(c-1)!} \frac{d}{d\rho} \left[\frac{(1-\rho^{k-c+1})}{(1-\rho)} \right]$$

$$= \frac{p_0 r_1^{c-1} \rho^2}{(c-1)!} \left[\frac{(1-\rho^{k-c+1}) - (1-\rho)(k-c+1)\rho^{k-c})}{(1-\rho)^2} \right]$$

We can now find the expected system length, L, as follows:

$$L_q = \sum_{n=c}^{k} (n-c) p_n$$

$$= \sum_{n=c}^{k} n p_n - c \sum_{n=c}^{k} p_n$$

$$= \sum_{n=0}^{k} n p_n - \sum_{n=0}^{c-1} n p_n - c \sum_{n=c}^{k} p_n$$

$$= L - \sum_{n=0}^{c-1} n p_n - c(1 - \sum_{n=0}^{c-1} p_n)$$

$$= L - \sum_{n=0}^{c-1} (n-c) p_n - c$$

Therefore, $L = L_q + c + \sum_{n=0}^{c-1} (n-c) p_n = L_q + c - p_0 \sum_{n=0}^{c-1} \frac{(c-n) r_1^n}{n!}$.

Having found L, we can now proceed to find the optimal allowable queue length and subsequently determine the optimal congestion toll in order to achieve that queue length.

The existing literature takes a common approach to arrive at the optimal queue length and congestion toll. Lippman and Stidham, (1977), Knudsen (1972), Yechiali (1971 and 1972), Edelson and Hilderand (1974) and Naor (1969) all use more or less the same technique. The general idea is to maximize some kind of welfare function by choosing an optimal allowable queue size. Then an optimal toll that induces the private agents to choose the optimal queue length is determined. Here, however, we solve a cost minimization, instead of a welfare maximization, problem. Cost minimization enables us to avoid the problem of estimating the direct benefits accruing to ships berthing at Port Klang. We will minimize the following cost function with respect to 'k' to get the optimal queue length.

$$E_k = bF + C_q L$$

$$= \lambda F\left[\frac{\lambda^k p_0}{(c-1)!\mu_1^{c-1}c^{k-c+1}\mu^{k-c+1}}\right]$$

$$+ C_q\left\{\frac{p_0 r_1^{c-1}\rho^2}{(c-1)!}\left[\frac{(1-\rho^{k-c+1})-(1-\rho)(k-c+1)\rho^{k-c}}{(1-\rho)^2}\right] + c - p_0\sum_{n=0}^{c-1}\frac{(c-n)r_1^n}{n!}\right\}$$

$$= \lambda F\left[\frac{\lambda^k\left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!}+\frac{r_1^{c-1}}{(c-1)!}\frac{\left(1-\rho^{k-c+1}\right)}{\left(1-\rho\right)}\right]^{-1}}{(c-1)!\,\mu_1^{c-1}c^{k-c+1}\mu^{k-c+1}}\right]$$

$$+ C_q\left\{\frac{\left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!}+\frac{r_1^{c-1}}{(c-1)!}\frac{(1-\rho^{k-c+1})}{(1-\rho)}\right]^{-1} r_1^{c-1}\rho^2}{(c-1)!}\left(\frac{(1-\rho^{k-c+1})-(1-\rho)(k-c+1)\rho^{k-c}}{(1-\rho)^2}\right)\right.$$

$$\left. + c - \left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!}+\frac{r_1^{c-1}}{(c-1)!}\frac{(1-\rho^{k-c+1})}{(1-\rho)}\right]^{-1}\sum_{n=0}^{c-1}\frac{(c-n)r_1^n}{n!}\right\}$$

where: b is the number of ships that are refused entry per unit time and is equal to λp_K

F is the penalty for loading/unloading at an alternative port.

C_q is the cost of waiting per unit time per ship.

The first term in the equation is the expected cost to the society per unit time for diverting traffic to an alternative port for a given choice k^*. The second term is the expected cost of waiting per unit time for a given choice of k^* for all ships arriving and joining the queue during the period. The optimal level of k (i.e. k^*) is determined by choosing that level of k that minimizes the function E_k.

The general strategy is for us to compare different levels of costs to the society by varying the level of maximum allowable queue length. The optimal level of 'k' is chosen when it is not possible to find another 'k' such that the cost is reduced. Recall, also that there is a level of 'k' i.e. k^p that will be chosen by the private agents. There are two counteracting forces that will give us the minimum point as we vary the level of allowable queue length. On the one hand, as we increase k, fewer ships will have to go to an alternative port thus reducing the overall cost of loading/unloading. On the other hand, the average waiting time for ships also gets longer.

In order to achieve the socially optimal queue length, the following inequalities have to hold: $E_k - E_{k+1} \leq 0$ and $E_{k-1} - E_k > 0$. The two inequalities tells us to continue increasing/decreasing the level of 'k' as long as the marginal social cost is lower/more than zero. The logic of the analysis can be explained as follows. Suppose for the sake of argument, we shorten the length of allowable queue from what is optimal. With a shorter queue length, more ships will be turned away, given the same arrival rate. On the one hand, total cost to the society will increase because ships that were turned away would have to incur a higher cost to transport the goods to or from its alternative destination. On the other hand, we have a counter effect to the total cost because with shorter line, every ship joining the queue will need to stay for a shorter period of time in port, resulting in lower waiting cost. The opposite is true, of course, when we lengthen the allowable queue. Fewer ships will be turned away, thus saving the overall cost of transportation. However, each ship that joins the queue will be expected to stay for a longer period of time. The optimal point is found when there is no advantage in either increasing or decreasing the allowable queue length.

The two inequalities, $E_{k-1} - E_k > 0$ and $E_k - E_{k+1} \le 0$, when expanded will take the following form:

$$E_{k^*-1} - E_{k^*} = \lambda F \left\{ T \left[\frac{\rho^{k^*-c}}{R + \frac{S(1-\rho^{k^*-c})}{(1-\rho)}} - \frac{\rho^{k^*-c+1}}{R + \frac{S(1-\rho^{k^*-c+1})}{(1-\rho)}} \right] \right\} +$$

$$+ C_q \left\{ \begin{array}{l} U \left[\dfrac{(1-\rho^{k^*-c}) - (1-\rho)(k^*-c)\rho^{k^*-c-1}}{R + \frac{S(1-\rho^{k^*-c})}{(1-\rho)}} - \dfrac{(1-\rho^{k^*-c+1}) - (1-\rho)(k^*-c+1)\rho^{k^*-c}}{R + \frac{S(1-\rho^{k^*-c+1})}{1-\rho}} \right] \\ - V \left[\dfrac{1}{R + \frac{S(1-\rho^{k^*-c})}{(1-\rho)}} - \dfrac{1}{R + \frac{S(1-\rho^{k^*-c+1})}{1-\rho}} \right] \end{array} \right\} > 0$$

$$E_{k^*} - E_{k^*+1} = \lambda F \left\{ T \left[\frac{\rho^{k^*-c+1}}{R + \frac{S(1-\rho^{k^*-c+1})}{1-\rho}} - \frac{\rho^{k^*-c+2}}{R + \frac{S(1-\rho^{k^*-c+2})}{1-\rho}} \right] \right\} +$$

$$+ C_q \left\{ \begin{array}{l} U \left[\dfrac{(1-\rho^{k^*-c+1}) - (1-\rho)(k^*-c+1)\rho^{k^*-c}}{R + \frac{S(1-\rho^{k^*-c+1})}{1-\rho}} - \dfrac{(1-\rho^{k^*-c+2}) - (1-\rho)(k^*-c+2)\rho^{k^*-c+1}}{R + \frac{S(1-\rho^{k^*-c+2})}{1-\rho}} \right] \\ - V \left[\dfrac{1}{R + \frac{S(1-\rho^{k^*-c+1})}{(1-\rho)}} - \dfrac{1}{R + \frac{S(1-\rho^{k^*-c+2})}{1-\rho}} \right] \end{array} \right\} \le 0$$

where $R=\sum_{n=0}^{c-1}\frac{r_1^n}{n!}$, $S=\frac{\rho r_1^{c-1}}{(c-1)!}$, $T=\frac{r_1^{c-1}}{(c-1)!}$, $U=\frac{r_1^{c-1}\rho^2}{(c-1)!(1-\rho)^2}$, $V=\sum_{n=0}^{c-1}(c-n)\frac{r_1^n}{n!}$

Concentrating on $E_{k^*-1}-E_{k^*}$ first, we can get the following expression after some manipulations.

$$E_{k^*-1}-E_{k^*} = \lambda F\{T[(1-\rho)\rho^{k^*-c}[R(1-\rho)+S(1-\rho^{k^*-c+1})]-(1-\rho)\rho^{k^*-c+1}[R(1-\rho)+S(1-\rho^{k^*-c})]]\}$$

$$+C_q\left\{\begin{matrix} U\begin{bmatrix}(1-\rho)[(1-\rho^{k^*-c})-(1-\rho)(k^*-c)\rho^{k^*-c-1}][R(1-\rho)+S(1-\rho^{k^*-c+1})] \\ -(1-\rho)[(1-\rho^{k^*-c+1})-(1-\rho)(k^*-c+1)\rho^{k^*-c}][R(1-\rho)+S(1-\rho^{k^*-c})]\end{bmatrix} \\ -V(1-\rho)[[R(1-\rho)+S(1-\rho^{k^*-c+1})]-[R(1-\rho)+S(1-\rho^{k^*-c})]]\end{matrix}\right\}>0$$

$$=\lambda F\{T(1-\rho)^2\rho^{k^*-c}[R(1-\rho)+S]\}$$
$$-C_q\{[V(1-\rho)^2\rho^{k^*-c}S]-U(1-\rho)^2\rho^{k^*-c-1}[(k^*-c)(2\rho R-R-\rho^2R-s+\rho S)+\rho S(1-\rho^{k^*-c})]\}>0$$

$$\Rightarrow \quad \lambda F\{T(1-\rho)^2\rho^{k^*-c}[R(1-\rho)+S]\}>$$
$$C_q\{[V(1-\rho)^2\rho^{k^*-c}S]-U(1-\rho)^2\rho^{k^*-c-1}[(k^*-c)(2\rho R-R-\rho^2R-s+\rho S)+\rho S(1-\rho^{k^*-c})]\}$$

$$\Rightarrow \quad \frac{Fc\mu}{C_q}>\frac{\{[V(1-\rho)^2\rho^{k^*-c}S]-U(1-\rho)^2\rho^{k^*-c-1}[(k^*-c)(2\rho R-R-\rho^2R-s+\rho S)+\rho S(1-\rho^{k^*-c})]\}}{\{T\rho(1-\rho)^2\rho^{k^*-c}[R(1-\rho)+S]\}}$$

Therefore,

$$\frac{Fc\mu}{C_q}>\frac{\{[VS]-U\rho^{-1}[(k^*-c)(-R)(1-\rho)^2+(k^*-c)(-s+\rho S)+\rho S(1-\rho^{k^*-c})]\}}{\{T\rho[R(1-\rho)+S]\}}$$

The other requirement to be satisfied in order to find the minimum point is $E_k - E_{k+1}\le 0$.

Using a similar procedure as applied above, this condition translates into:

$$\frac{Fc\mu}{C_q}\le\frac{\{[VS]-U\rho^{-1}[(k-c)(-R)(1-\rho)^2+(k-c)(-s+\rho S)+\rho S(1-\rho^{k-c})]\}}{\{T\rho[R(1-\rho)+S]\}}$$

Combining both inequalities, we get:

$$\frac{\left\{[VS]-U\rho^{-1}\begin{bmatrix}((k+1)-c)(-R)(1-\rho)^2\\+((k+1)-c)(-s+\rho S)\\+\rho S(1-\rho^{(k+1)-c})\end{bmatrix}\right\}}{\{T\rho[R(1-\rho)+S]\}}\geq\frac{Fc\mu}{C_q}>\frac{\left\{[VS]-U\rho^{-1}\begin{bmatrix}(k-c)(-R)(1-\rho)^2\\+(k-c)(-s+\rho S)\\+\rho S(1-\rho^{k-c})\end{bmatrix}\right\}}{\{T\rho[R(1-\rho)+S]\}}$$

Recall that $\frac{Fc\mu}{C_q}=Z_p$ is the value used to the determine the privately optimal level of queue length. To show that the privately optimal system size is generally smaller than the socially optimal one, we will investigate the function:

$$Z_p=\frac{\left\{[VS]-U\rho^{-1}\left[(z^*-c)(-R)(1-\rho)^2+(z^*-c)(-s+\rho S)+\rho S(1-\rho^{z^*-c})\right]\right\}}{\{T\rho[R(1-\rho)+S]\}}$$

If we can show that Z_p is always increasing in Z^*, than the integers k^* and k^*+1 will always satisfy the above inequality. As a consequent we can conclude that k^* is the biggest integer not greater than Z^*. Notice, however, that satisfying the inequality does not imply that the biggest difference between k^* and k^p is one, just because it appears that Z^* is sandwiched between k^* and k^*+1. We shall demonstrate in TABLE 1 below the relationship between k^* and k^p for different levels of ρ.

To prove Z_p is always increasing in Z^*, we differentiate the function with respect to Z^* and get the following:

$$\frac{\partial Z_p}{\partial Z^*}=\frac{-U\left[(\rho S-R(1-\rho)^2-S-\rho^{Z^*-c+1}S\log\rho\right]}{\{T\rho^2[R(1-\rho)+S]\}}$$

We can prove that the derivative is greater than zero. The proof is fairly cumbersome, however, so we relegate it to Appendix 2, Section A.

We need now to show that $Z_p \geq Z^*$ and this can be done by showing that Z_p - $Z^* \geq 0$. Again the proof is not straight forward and we show it in Appendix 2, Section B. Certainly, the fact that $Z_p - Z^* \geq 0$, does not always imply k^* is less than k^p since it is possible to find both Z_p and Z^* falling between the same integers. However, the most interesting cases for congestion studies deal with periods of heavy traffic in which ρ is high. At the same time the fact that k^* is generally lower than k^p, means that the port authority ought to limit queue size through, for example, imposing some kind of congestion toll to provide enough disincentive for private operators to call at a port under this circumstances. This is indeed the time when the difference between Z_p and Z^* is the highest. Table 1 below demonstrates some of the assertions that we have made regarding the relationship between the variables. It should also serve to give us a general idea about the relationship between the variables, in particular k^* and k^p.

What can we learn from the table above? First, the privately optimal strategy k^p is not affected by the degree of port utilization as measured by ρ. This is represented by a constant k^p for each column, even though ρ increases as we go down each column. The reason is fairly straight forward. When considering whether or not to join a queue, a ship operator will only look at the existing queue size and will not consider whether his decision will affect the waiting time for ships arriving after him. In doing so, his strategy, k^p, will not be affected by the intensity of port utilization because what matters is only the size of the queue when he arrives, which in turn affects the trade-off between queuing and penalty cost. This observation certainly has the economic intuition that

Table 1

Socially Optimal Queue Length (k^*) as a function of Z_p and ρ
with the assumption c=2 and a=0.8 where a is the ratio μ/μ_1

ρ	Z_p=2.5 or k^p=2	Z_p=3.5 or k^p=3	Z_p=4.5 or k^p=4	Z_p=5.5 or k^p=5	Z_p=6.5 or k^p=6	Z_p=10.5 or k^p=10	Z_p=15.5 or k^p=15
0.1	2	3	4	5	6	9	14
0.2	2	3	3	4	5	8	12
0.3	2	3	3	4	5	7	11
0.5	2	2	3	3	4	6	8
0.7	2	2	3	3	4	6	8
1.0	2	2	2	3	3	4	5
1.5	2	2	2	2	3	3	4
2.0	2	2	2	2	2	3	3
2.5	2	2	2	2	2	3	3
3.0	2	2	2	2	2	3	3

private agents do not generally consider the externality they impose on others when making their decisions.

The second observation is that as we go down towards the right hand corner we find an increasing gap between k^p and k^*. To explain this pattern we need to state a few basic facts about the table. Going down the table means higher and higher ρ . Higher ρ reflects higher arrival rate relative to the service rate. In other words, the port gets relatively more congested as we go down the table. Going across the table to the right corresponds to increasing k^p. It means that the penalty cost is getting larger relative to the queuing cost. When we combine these two observations, we will get the explanation for the pattern that is evident in the table. The explanation is as follows:

As we move down towards the lower right hand corner of the table, the congestion externality that port users impose on one another gets larger and larger for two reasons. First, as the penalty becomes relatively bigger, the private strategy calls for ships to join longer and longer queues. As queues get longer, each additional arrival will congest more ships that arrives later. In total the sum of negative externalities increases as queues lenghten. However, each individual ship operator only compares its private waiting costs against the higher penalty. Each ship operator ignores the greater and greater negative externality he imposes on the others as the queue gets longer and longer. Now the optimal queue length ought to be increased also as the penalty becomes bigger because it is becoming more expensive to turn away a ship. However, the socially optimal allowable queue size will not increase by as much that called for by the private strategy. Thus there is an increasing gap as we move to the right of the table.

Now, as we move down the table along any column, the port essentially becomes more and more congested. Suppose that this greater congestion is due to an increasing rate of arrivals. It is true that the private strategy will not change even if the arrival rate increases because to ship operators the relative cost between waiting and paying the penalty has not changed. However, with greater arrivals, (even with an unchanged privately determined queue length) the expected queue length will still be larger. Queues will be more fully occupied, thus increasing the externality that each ship imposes on the others. Again this will result in a bigger gap between the private and social queue length because greater and greater externality imposed will eventually leads to the prescription of shorter and shorter socially optimal queue.

The next section derives the optimal toll charges as a way of inducing the ship operators to choose a privately optimal strategy that is also socially optimal.

Optimal Toll Charges

Toll charges are needed in order to induce private operators to carry out socially optimal strategy. As we explained above, achieving social optimality requires us to shorten queue size as the port gets more congested and the penalty cost (i.e. the cost of turning ships away) gets lower. One way of doing this is through the imposition of congestion toll. (Another way is , of course, to do it administratively by just instructing ship not to join the queue when the size reaches a predetermined limit.) In this section we derive the rule for arriving at the optimal toll charges.

Notice first that to a private operator, the gain from queuing gets smaller and smaller as the queue size gets larger until it can even become negative. The reason is fairly evident since the penalty for balking is fixed but the waiting cost increases as the queue gets longer. Therefore, in order to induce private operators to choose k^* instead of k^p, we must charge each arriving ship a fee (which we call the congestion toll), that is large enough to discourage any ship to queue beyond k^*. However, this fee must be small enough, at the same time, to make any ship to want to queue when it finds a queue size of k^*-1 when it arrives in port.

As far as our model is concern, that congestion toll, t, is given by the following inequality:

$$F - C_q\left(\frac{k^*+1}{c\mu}\right) < t \leq F - C_q\left(\frac{k^*}{c\mu}\right)$$

Now, the left hand side of the inequality is the gain that is to be made by a ship that arrives and join the queue when the number of ships waiting is k^*. Now, as the inequality indicated, we want the toll charge to be bigger than this amount so that no ship will join the queue when the number of waiting ship is already k^*. On the right hand side of the inequality we have the gain for a private operator that arrives and join the queue when the number of waiting ship is k^*-1., When the number of ships in the system is k^*-1, we want the ship to join the queue since the socially optimal queue length is k^*. This necessarily means that our toll must not exceed this amount because otherwise the ship will balk even before the optimal queue size is reached.

It is evident that the rule that we stated above calls for higher and higher toll charges if the penalty gets higher and/or the waiting cost gets lower. The reasoning is fairly similar to that which we give in explaining TABLE 1 earlier. Higher penalty and lower waiting cost encourage private ship operators to join larger queues while the greater externality they impose on the others. This call for shorter queue size for social optimality, which in turn, is achieved by levying higher toll charges.

Having determined the optimal toll, we are now ready to move to the final section of this chapter which is to find the optimal investment pattern.

Optimal Investment

Optimal investment was not considered in those papers we listed earlier. However, we devote this section towards determining the optimal investment pattern in port. At the same time, we differ from past researches that have dealt with the issue of optimal investment (like Wanhill (1974) and Weille (1974)), because they calculated the optimal pattern without incorporating optimal toll and queuing behavior in their model. In their model, the gain from building a new berth is then computed as the savings made in lowering the queuing cost for all ships because they assume that no ship ever balks.

We calculate the gain from investing in new berth by calculating the cost and gain after taking into consideration that optimal tolls are imposed to regulate queue size before and after the construction of the new berth. The gain from a marginal investment in our model will come from two sources. First, a gain is to be made through lower waiting cost for each ship because the berths can now turn ships around at a faster rate, given the arrival rate and the allowable queue length. The second source of gain will come from savings made in lower overall penalty cost because of fewer ships balking as more ships can be loaded/unloaded for a given period of time.

Before we go ahead with deriving the optimal investment pattern, it may be worthwhile to see how the relationship between k^* and k^p changes as we increase the number of berths in a subsystem. We re-compute Table 1, but this time we will work on the assumption that there is one additional berth in the sub-system. For easy comparison, we provide the corresponding values for k^* under one fewer berth in brackets. Z_p increases by fifty percent when we adjust the variable 'c' to reflect the increase in the number of berths. We continue to make the assumption that the ratio μ/μ_1 equals 0.8.

We omit the same general observations that we made for the case of a two-berths system. Instead, we first observe that the privately optimal queue size increases in all cases. The reason is fairly evident. With faster service, it is now cheaper for ship operators to queue no matter what is the queue size, hence, their strategy will involve bigger queue size. Second, the optimal allowable queue size also increases in all cases. The explanation is as follows:

With higher berth servicing capacity, the expected queue size and hence waiting time will be smaller. As a result, the congestion externality will be lower because each ship will be expected to congest fewer other ships and also for a shorter period of time. At the same time, the penalty for loading/unloading at an alternative port has not changed. A lower externality now upsets the previous optimal trade-off between the cost of queuing and the penalty cost and in fact moves towards the direction of larger allowable queue size. This explains the increase in optimal queue size when we add another berth to the subsystem. To arrive at the optimal investment pattern we compare the discounted

Table 2

Socially Optimal Queue Length (k^*) as a function of Z_p and ρ with the assumption c=3 and a=0.8 where a is the ratio μ/μ_1. Corresponding values for c=2 is given in brackets.

ρ	Z_p=3.75 or k^p=3	Z_p=5.25 or k^p=5	Z_p=6.75 or k^p=6	Z_p=8.25 or k^p=8	Z_p=9.75 or k^p=9	Z_p=15.75 or k^p=15	Z_p=23.25 or k^p=23
0.1	3(2)	5(3)	6(4)	7(5)	9(6)	14(9)	21(14)
0.2	3(2)	4(3)	5(3)	7(4)	8(5)	13(8)	19(12)
0.3	3(2)	4(3)	5(3)	6(4)	7(5)	11(7)	17(11)
0.5	3(2)	4(2)	4(3)	5(3)	6(4)	9(6)	13(8)
0.7	3(2)	3(2)	4(3)	4(3)	5(4)	7(6)	9(8)
1.0	3(2)	3(2)	3(2)	4(3)	4(3)	5(4)	6(5)
1.5	3(2)	3(2)	3(2)	3(2)	4(3)	4(3)	5(4)
2.0	3(2)	3(2)	3(2)	3(2)	3(2)	4(3)	4(3)
2.5	3(2)	3(2)	3(2)	3(2)	3(2)	4(3)	4(3)
3.0	3(2)	3(2)	3(2)	3(2)	3(2)	4(3)	4(3)

benefits against the discounted costs of adding a new berth. The economic net present value for the marginal investment can be written as:

$$NB_t = \sum_{t=0}^{n} \frac{1}{(1+r_t)^t} \{(W_t + F_t) - (K_t + M_t)\}$$

where W_t is the gain/loss in waiting time cost

F_t is the savings in penalty cost through fewer ships having to load and unload elsewhere.

K_t is the capital outlay

M_t is the maintenance cost

We assume here that at the end of the economic life of the berth, n, it has a salvage value of zero, which is a reasonable assumption for a specialized investment. We can further define

$$W_t = (L_t^c - L_t^{c+1})C_q$$

and $F_t = F\{\lambda(1-p_*^{c+1}) - \lambda(1-p_*^c)\} = F\lambda(p_*^c - p_*^{c+1})$. The superscripts '*c*' and '*c+1*' for variable *L* refers to the optimal values for *L* when the number of berths is '*c*' and '*c+1*' respectively. The subscript (*) on *p* refers to the probability of having the system size equal to the allowable queue length when the number of berth are '*c*' and '*c+1*'.

Therefore we can rewrite the net economic present value as:

$$NB_t = \sum_{t=0}^{n} \frac{1}{(1+r_t)} \{[(L_t^c - L_t^{c+1})C_q] + [F\,\lambda(p_*^c - p_*^{c+1})] - (K_t + M_t)\}$$

If the net present value is greater than zero, the investment should be made.

Of course, the analysis can be extended if we have growth in the number of ship arrivals. In fact a different rate of growth will result in different optimal timing of investment for new berths. We can incorporate growth in ship arrivals by having different arrival rates, λ, from one time period to another. It also appears that the welfare measures used in the formula above are not entirely national gains or loss because probably most ships that called at Port Klang that incur the cost of queuing and penalty are not Malaysian ships. However, the counter argument is that ultimately these costs will be reflected in shipping charges or the price of goods paid by Malaysian nationals.

Conclusion

In this chapter we present the state dependent multi-server queuing model where we derive several expressions for calculating the optimal congestion toll and investment pattern. These expressions will be used again in the empirical part of our study.

The model is an extension of the standard queuing models where we relax the assumption of constant service time. Using the model we determine the socially optimal allowable queue length by minimizing the sum of queuing and penalty costs to the society. Privately optimal strategy is also derived in this chapter. We demonstrate that the strategy of the private ship operators always results in longer queue compared to the socially optimal level. This finding implies that, for social optimum, we need to device a mechanism to limit queue size. One such mechanism is the imposition of a suitable toll schedule to arriving ships. In Chapter 4 we will use the model of this chapter to empirically determine the optimal congestion toll charges and investment timings for different subsystems at Port Klang.

Chapter 4

Empirical Analysis

There are two major goals in this chapter. First, we test the assumption of independence between service and congestion rate. Second, we identify various other factors that significantly influence the amount of time that ships stay in port for the purpose of loading/unloading their cargo (service time). Both goals are achieved through the same regression analysis. We begin by describing the data, followed by a discussion of the regression model. In the last section we present and discuss the results of the regression.

The Data

The data cover all twenty-six berths (numbered 1 to 25 and 7A) serving different types of cargoes and ships at Port Klang. For the purpose of our analysis, the port is divided into seven sub-systems. Each sub-system is naturally divided into the kind of cargo and ship it serves. The berths in each subsystem are also physically clustered together. Hence, berths 8 to 11 and 19 to 21 are treated as two sub-systems (even though they both serve container ships) because they are physically apart. Institutionally, they are also run by two distinct corporations with separate operations and no sharing of berth facilities. The divisions and descriptions of the berths are shown in Table 3.

Table 3

Descriptions of Cargoes Handled by Different Subsystems of Berths

Berth Number	**Description**
1 and 2	Liquid Bulk cargo (e.g. palm oil, latex and petroleum)
5,6,7 and 7a	Break Bulk cargo (e.g. bagged rice and sugar, palm oil in drums, palm kernel and timber)
8,9,10 and 11	Container cargo
12,13,14,15,16,17 and 18	Break Bulk cargo (e.g. bagged rice and sugar, palm oil in drums, palm kernel and timber)
19,20 and 21	Container cargo
22 and 23	Liquid Bulk cargo (e.g. palm oil and petroleum)
24 and 25	Dry Bulk Cargo (e.g. chemical and wheat)

We have data for a period of twenty-five months (September 1993 to September 1995), covering each ship that called at Port Klang during the period. All of the data used in this research were taken from the port operator's data base. The data were collected by the port management using a two-step procedure. Initially, a time sheet is filled out for each ship calling at Port Klang by several employees in charge of berth operations. These time sheets are then collected on a weekly basis by the port planning division either to be kept for further analysis or to be keyed into a computer spreadsheets. The time sheets and spreadsheets contain information on ships' and cargoes' characteristics. For each ship that called at Port Klang we have information on the time it arrived and left the dock (service time) as well as the amount and type of cargo it carried.

Using the information on dock arrivals and departures, we derive two occupancy variables. These variables are the subsystem and port wide occupancy. We have been careful in constructing these variables in order to avoid a mechanical relationship with the service time. The exact formula employed and the rationale are described in the next section.

For the non-container berths the cargo amount is measured in tonnes while for the container terminals it is the number of twenty-foot-equivalent container boxes (TEU). TEU is a standard industry measure for container throughput.

We also have a detailed description of the cargo type carried by each ship for all terminals. For example in the case of berths 1 and 2, the cargo categories are described as palm oil, petroleum, latex or chemical. Only in the case of the breakbulk terminals (berths 5-7A and 12-18) where each ship tends to carry a greater variety of cargo, the description are less detailed. For example ships carrying different types of cargo are just described as 'general' for these berths.

In terms of data quality, we found that there was no data cross checking or auditing mechanism implemented by the port operators as a measure to ensure the accuracy of their database. This means errors remain uncorrected once entered into the database. These unavoidable recording errors are always a problem with the data, but we have no way of assessing its severity. However, with the knowledge that the data were systematically recorded on site without delay, there is no reason why the degree of error is higher than what is normally expected. At the same time we do not expect systematic errors that may bias our results in a certain direction.

There are probably data on several other variables that can be useful in predicting service time but were, unfortunately, unavailable. These include data on the intensity of port side loading/unloading inputs per unit time and a measure of ship side effectiveness and quantity of loading and unloading inputs. The likely bias on the estimated regression coefficients due to the lack of data on these variables will be discussed below.

The Regression Model

While the primary purpose of the regression analysis is to test the independence hypothesis, we determine other factors that may systematically influence service time. Towards this end, we specify the regression as follows:

$$\begin{aligned}\text{Log Service} = {} & \beta_0 + \beta_1 \text{Congestion} + \beta_2 \text{Congestion Squared} + \beta_3 \text{Port Wide Congestion} \\ & + \beta_4 \text{Port Wide Congestion Squared} + \beta_5 \text{Log Throughput} \\ & + \beta_{6,1} \text{Cargo Dummy 1} + \beta_{6,2} \text{Cargo Dummy 2} \ldots \\ & + \beta_{7,1} (\text{Cargo Dummy 1})^* \text{Congestion} + \beta_{7,2} (\text{Cargo Dummy 2})^* \text{Congestion} \\ & + \beta_{8,1} (\text{Cargo Dummy 1})^* \text{Congestion Squared} \\ & + \beta_{8,2} (\text{Cargo Dummy 1})^* \text{Congestion Squared} \ldots \\ & + \beta_{9,1} (\text{Cargo Dummy 1})^* \text{Log Throughput} \\ & + \beta_{9,2} (\text{Cargo Dummy 1})^* \text{Log Throughput}\end{aligned}$$

We regress an individual ship's log of service time against a measure of berth occupancy, port wide congestion, port wide congestion squared, log throughput, cargo dummies and dummies interacted with congestion, congestion squared and log throughput. We provide the definition of each variable used in the regressions in Table 4. The decision to include the variables are motivated by theory while the choice of functional form is partially influenced by bivariate plots presented in Appendix 3.

Choosing a suitable definition for the occupancy rate which is also amenable to regression analysis is not as straight forward as it first appears. The standard definition of occupancy rate as regularly used in the literature as well as by many port authorities as one of port performance indicators, has as the numerator, the sum of the time that the various berths are occupied. Since the service rate also has the same variable as the numerator, the result of the regression will be biased towards confirming our hypothesis.

Instead of using the standard definition, we measure the congestion level and service time experienced by each individual ship. This measure is not biased towards confirming our hypothesis because of the following reason. The variable 's' (service time) will be defined as the duration that a ship spent in port while in service. Let c_j be the length of time for each 'j' other ships that are present simultaneously with the ship under consideration. If there are n berths in the sub-system, then j=2,3....n. The congestion index, 'c' is defined as follows:

$$c = (a_2c_2 + a_3c_3 \; .... + a_n c_n) / s$$

where $a_2, a_3 ......\ a_n$ are the weights that are put on the different time periods that 2, 3.......n ships are present simultaneously. For illustrative purposes, suppose that the distribution of ships are as follows for a particular day in a two-berth system:

Berth 1 ________________ Ship x

Berth 2 __________________________________ Ship y

00:00 8:00 16:00 24:00 hours

Ship 'x' arrived at berth 1 at 00:00 hrs and stayed for 8 hrs. Ship 'y' stayed at berth 2 for 16 hrs. Our congestion index for ship x in this example is {(0 x 0)+(1 x 8)}/8 = 1.00. The index assigned to ship y is {(0 x 8) + (1 x 8)}/16 = 0.50. The indices were found using a weight of zero if no other ship was present and one if there is one other ship present simultaneously. This weighting certainly makes sense since when there is no other ship present, the ship under consideration does not experience any congestion. Maximum congestion happens, in a two-berth subsystem, when there are two ships berthing simultaneously, thus the weight in this case, equals to one.

The new congestion index has a strong economic justification, too. The index is always higher whenever we have more ships in port *simultaneously*. The standard occupancy rate measurement, on the other hand, does not guarantee that we always have a higher index when the port is more congested. For example, the standard definition of occupancy ($s_1 + s_2 + s_3$............s_n)/(2x24xn) does not differentiate between the following two scenarios.

Scenario 1

Berth 1 ______________ Ship x

Berth 2 ______________ Ship y

00:00 8:00 16:00 24:00 hours

Scenario 2

Berth 1 ______________ Ship x

Berth 2 ______________ Ship y

00:00 8:00 16:00 24:00 hours

According to the standard definition, the occupancy rate in both scenarios is (8+8)/48=.333. However, the new congestion index gives us c={(1 x8) + (1+8)}/8=1 for the first scenario and {(0 x 8) + (1 x 0)}/8=0 for the second scenario for each ship. Clearly, the second definition is better for our purpose because the first scenario is more congested as the port resources will have to cater for two ships simultaneously. The port in the second scenario is free of any congestion since each ship is worked on one at a time.

Next, notice that there are two occupancy/congestion variables; one for a measure for within the sub-system (congestion) and the second for the whole port (port wide congestion). The congestion variables are expected to influence service time since we hypothesize that some resources are shared across berths and the loading/unloading process becomes more cumbersome as the port apron and storage area becomes congested. We add a quadratic term for the congestion variables because the relationship between congestion and service time is not expected to be linear. In addition preliminary bivariate plots between service time and congestion indicate service time initially increases but later falls as congestion rises. These plots were presented in Figure 4 of Appendix 3. A log transform is chosen for service time because it allows for a more rapid decline in service time when (subsystem) congestion rises as suggested by the preliminary plots. Note that for berths 5-7A and 12-18 there are very few points where the occupancy rate falls below the 35% level. We therefore have to be very cautious when interpreting the regression for occupancy rate below this level.

The bivariate plots for the port wide congestion variable (Figure 5, Appendix 3) send a mixed signal as to the probable relationship between port wide congestion and service time. For some subsystems port wide congestion appears not to influence service time while for others there exists a strong quadratic relationship. Notice also there are very few data points below the 30% occupancy level for all berths.

The cargo dummies are included because of the likelihood for the expected service time to be affected by cargo types. For example, ships carrying mixed cargo are expected to have greater expected service time than ships transporting standardized cargo, holding other factors constant.

Cargo throughput should also have an impact on service time. A larger cargo is expected to take longer to load/unload. Of particular interest is the choice of a log transform on this variable. Once service time is log transformed we need also a log transformation on the throughput variable in the regression to allow for the strong likelihood of service time increasing proportionally slower than cargo size. Note however that this specification still allows for constant returns to scale in cargo size. In any case, the bivariate plots between service time and throughput in Figure 6 of Appendix 3 appear to support the increasing returns to scale hypothesis.

Our hypothesis of dependence between service and occupancy rate cannot be rejected if β_1 and β_2 are jointly significantly different from zero. The hypothesis that ships also congest other ships that are not in their sub-system is tested by finding out whether β_3 and β_4 are jointly significantly different from zero. The coefficients on the interactions between dummies and congestion or throughput variables are used to test whether service times for ships carrying different cargoes are affected differently by congestion and throughput. We also test the possibility that the variance of service time may be a function of one or several of the independent variables by exploiting the high possibility of heteroscedastic variances in our cross sectional data.

There are possible other variables that can have some impact on service time for which we unfortunately have no data. We therefore only discuss their likely effect on the estimated coefficients of the current model. The first variable is a measure of intensity of port side loading/unloading inputs per unit time per ship. The amount of loading/unloading inputs devoted to a ship per unit time should have a negative influence on the average service time. If we further assume that input intensity is positively correlated with throughput (because it is likely that the port operator assigns more loading/unloading inputs per unit time to ships carrying bigger amount of cargo), the estimated coefficient for the log throughput variable is biased downwards if the input

TABLE 4

Definitions of variables used in the regression.

Variable	**Definition**
Log Service	Log of the service time for each individual ship i.e. log of the departure minus arrival times in hours.
Congestion	$(a_2c_2 + a_3c_3 \ldots\ldots + a_n c_n) / s$ where a_i is the duration when there are 'i' ships present simultaneously and c_i is the proportion of the system occupied when there are 'i' ships in the subsystem. 's' is the duration spent at the quay side for loading/unloading.
Log Throughput	Total cargo loaded/unloaded during the time that a ship was in port in unit of 1000 tonnes. In the case of container berths the measure is in TEUS (Twenty-foot Equivalent Unit) which is the standard industry measure for container throughput.
Port Wide Congestion	The overall port wide congestion during the time that a ship concerned is at berth. This variable is calculated just like the congestion variable above, i.e. $(a_2c_2 + a_3c_3 \ldots\ldots + a_n c_n) / s$, except that a_i is the duration when there are i ships present simultaneously and c_i is the proportion of the system occupied when there are 'i' ships *in the port excluding the sub-system under consideration.*
Cargo Dummy	A variable (either 1 or 0) representing the various types of cargo like wheat, chemical, palm oil and iron and steel.

intensity variable is omitted. The coefficient is biased downwards because it includes the negative impact of the intensity variable on service time.

The omission of the input intensity variable also bias the estimate of the congestion coefficient downwards. Just as in the case of log throughput, the coefficient on congestion variable captures the negative influence of the intensity variable on service time. In a way this possibility works in favor of our hypothesis since it tends to give a more conservative estimate of the coefficient on the congestion variable.

The other probable omitted variable is a measure of the effectiveness of the ship-side loading/unloading inputs and the design of cargo hull which can possibly be proxied by the age of a ship. We speculate that newer ships tend to have a more modern loading/unloading facility as well as hull design that facilitate the process of loading/unloading. Other things equal, newer ship can be loaded faster than an older vessel because ship-side loading facility is more effective and better designed. If it is the case that the newer vessels tend to carry a greater amount of cargo because they are bigger, the estimate of the log throughput coefficient is biased downwards.

Finally, for all sub-systems, we check for heteroscedasticity using the Bruesch-Pagan test. If the test indicates heteroscedasticity for a sub-system, we run a Weighted Least Squares (WLS) by regressing the squared residuals of the ordinary least squares regression (OLS) against one or several of the independent variables. Our analysis then proceeds using the WLS estimates.

The next section describes the results of the regressions.

The Results

Our discussion of the results, takes the following approach. We first briefly describe the way the regression results are presented. We then discuss the results in general terms, paying particular attention to additional information that may be peculiar to a sub-system. The regression results are summarized in Tables 5 to 11.

The first column lists the variables for the regression. The second, third, fourth and fifth columns provide the corresponding OLS and WLS coefficients and t statistics. The final two columns give us the F test statistics for the joint test of significance as well as the level at which these statistics are significant. The explanatory note 1 (Additional F Tests) gives further results for the joint tests of significance for the listed variables. Results for the Bruesch-Pagan tests are presented at the bottom of the tables. The procedures of the Bruesch-Pagan tests are presented in Appendix 4.

Discussions of the results

We performed two regressions, one OLS and the other WLS, because the statistics for the Bruesch-Pagan test (a check for heteroscedasticity) are significant for all berths, which is expected for a cross section data like ours. Please refer to Appendix 4 for test results. The WLS procedure is implemented to get a more efficient estimates of the coefficients.

Table 5

Summary of Regression Results For Berths 1 & 2

N=623

Variables	**OLS**		**WLS**		**F-tests: WLS**	
	Coeff.	t-stats	Coeff.	t-stats	Ratio	P value
Constant	2.50	6.48	2.50	8.56		
Congestion	2.49	5.35	2.51	3.13		
Congestion squared	-2.68	-6.00	-2.66	-3.45		
Port Wide Congestion	0.26	0.23	0.27	0.39		
Port Wide Congestion squared	-0.14	-0.17	-0.22	-0.43		
Ln Throughput	0.27	6.38	0.30	3.79		
Cargo Dummies					6.45	0.000
Palmoil	-0.50	-4.47	-0.48	-2.65		
Petroleum	-0.23	-2.03	-0.23	-1.31		
Latex	-0.17	-0.54	-0.20	-0.48		
Chemical	-0.23	-1.01	-0.20	-0.61		
Congestion - Interacted with dummies					3.63	0.010
Palmoil-Congestion	-1.91	-3.74	-1.88	-2.26		
Petroleum-Congestion	-2.31	-4.33	-2.40	-2.93		
Latex-Congestion	-1.01	-0.81	-0.76	-0.47		
Chemical-Congestion	0.57	0.37	0.66	0.31		
Congestion squared - Interacted with dummies					3.85	0.000
Palmoil-Congestion Squared	2.08	4.26	2.03	2.54		
Petroleuml-Congestion Squared	2.46	4.78	2.51	3.19		
Latex-Congestion Squared	1.46	1.39	1.22	0.89		
Chemical-Congestion Squared	-0.06	-0.04	-0.20	-0.10		
Log Throughput - Interacted with dummies					13.28	0.000
Palmoil-Ln Throughput	0.22	4.68	0.20	2.52		
Petroleuml- Ln Throughput	-0.09	-1.62	-0.05	-0.56		
Latex- Ln Throughput	0.17	1.82	0.13	1.01		
Chemical- Ln Throughput	0.30	2.41	0.27	1.53		

Notes:

1. Additional F Tests	Ratio	P value
a. Port Wide Congestion & Port Wide Congestion squared	0.18	0.840
b. Congestion & Congestion squared	6.17	0.000
c. Congestion and its interactions	4.35	0.000
d. Congestion squared and its interactions	5.25	0.000
e. Both Congestion,Congestion squared and Interactions	3.05	0.000
f. Log Throughput and its interaction	59.8	0.000

Table 5 (cont'd)

2. Signs for cargo dummies are relative to mix cargo.

3. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
1532.73	7	0.99

Table 6

Summary of Regression Results For Berths 5-7A
N=609

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests: WLS** Ratio	P value
Constant	2.83	8.90	2.75	10.38		
Congestion	1.27	1.94	1.40	2.73		
Congestion squared	-0.62	-1.37	-0.70	-1.94		
Ln Throughput	0.50	17.43	0.50	18.93		
Port Wide Congestion	1.57	1.78	1.50	1.88		
Port Wide Congestion Squared	-1.47	-2.07	-1.34	-2.08		
Cargo Dummy						
Standard cargo	-0.76	-15.18	-0.80	-15.07		
Congestion Interaction						
Standard Cargo-Congestion	1.07	0.84	1.63	0.80		
Congestion squared Interaction						
Standard Cargo-Congestion Squared	-0.97	-1.07	-1.40	-0.99		
Log Throughput - Interaction						
Standard Cargo-Log Throughput	-0.05	-0.98	-0.20	-2.45		

Notes:

1. Additional F Tests	Ratio	P value
a. Port Wide Congestion & Port Wide Congestion squared	2.66	0.071
b. Congestion & Congestion squared	11.81	0.000
c. Congestion and its interactions	4.86	0.008
d. Congestion squared and its interactions	3.02	0.050
e. Both Congestion,Congestion squared and Interactions	6.48	0.000
f. Log Throughput and its interaction	192.80	0.000

2. Signs for cargo dummies are relative to general breakbulk cargo

3. Std cargo refers to ships that carry homogeneous cargo.

4. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
532.67	3	0.07

Table 7

Summary of Regression Results For Berths 8-11

N=1595

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests: WLS** Ratio	P value
Constant	-0.39	-3.99	-0.30	-2.70		
Congestion	0.36	2.21	0.31	1.72		
Congestion Squared	-0.12	-1.01	-0.04	-0.30		
Port Wide Congestion	-0.35	-1.36	-1.27	-5.20		
Port Wide Congestion squared	0.24	1.16	1.19	6.34		
Ln TEUS	0.47	52.42	0.48	42.57		

Notes:

1. F Tests

	Ratio	**P value**
a. Congestion & Congestion squared	29.78	0.000
b. Port Wide Congestion & Port Wide Congestion squared	33.08	0.000

2. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
1909.65	1	0.0004

Table 8

Summary of Regression Results For Berths 12 & 18
N=1213

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests** Ratio	**WLS** P value
Constant	0.04	0.03	-0.03	-0.03		
Congestion	7.32	2.69	7.43	2.61		
Congestion squared	-4.72	-2.81	-4.79	-2.72		
Port Wide Congestion	2.06	2.10	2.01	2.48		
Port Wide Congestion squared	-1.85	-2.15	-1.67	-2.35		
Ln Throughput	0.62	24.44	0.57	22.70		
Cargo Dummies					1.93	0.062
Vehicle	0.94	0.28	2.02	0.96		
Iron and Steel	-0.07	-0.04	0.07	0.04		
Chemical	3.60	2.55	3.73	2.87		
Palm Kernel	-0.77	-0.04	2.18	0.12		
Timber	2.85	0.29	2.22	0.42		
Sugar	-30.00	-1.49	-28.34	-1.16		
General	1.83	1.47	1.87	1.43		
Congestion - Interacted with dummies					3.63	0.006
Vehicle-Congestion	-4.58	-0.56	-8.00	-1.55		
Iron and Steel-Congestion	-2.02	-0.49	-2.27	-0.48		
Chemical-Congestion	-8.12	-1.96	-8.53	-2.23		
Palm Kernel-Congestion	0.90	0.02	-5.52	-0.14		
Timber-Congestion	-6.88	-0.31	-5.60	-0.47		
Sugar-Congestion	65.23	1.42	61.40	1.10		
General-Congestion	-4.83	-1.52	-4.70	-1.40		
Congestion squared - Interacted with dummies					0.96	0.456
Vehicle-Congestion Squared	3.36	0.68	5.67	1.82		
Iron and Steel-Congestion Squared	1.62	0.64	1.78	0.61		
Chemical-Congestion Squared	5.01	1.68	5.28	1.91		
Palm Kernel-Congestion Squared	-0.79	-0.03	2.77	0.12		
Timber-Congestion Squared	4.53	0.36	3.88	0.57		
Sugar-Congestion Squared	-35.33	-1.37	-33.21	-1.06		
General-Congestion Squared	3.25	1.60	3.09	1.43		
Log Throughput - Interacted with dummies					1.06	0.390
Vehicle-Log Throughput	0.02	0.34	-0.02	-0.54		
Iron and Steel-Log Throughput	-0.11	-2.76	-0.10	-2.11		
Chemical-Log Throughput	-0.10	-0.70	-0.04	-0.32		
Palm Kernel-Log Throughput	0.00	-0.34	0.00	-0.05		
Timber-Log Throughput	0.00	-0.02	0.05	1.32		
Sugar-Log Throughput	0.08	0.41	0.16	0.64		
General-Log Throughput	-0.12	-2.58	-0.14	-3.09		

Table 8 (cont'd)

Notes:

	Ratio	P value
1. F Tests		
a. Port Wide Congestion & Port Wide Congestion squared	3.30	0.037
b. Congestion & Congestion squared	4.08	0.017
c. Congestion and its interactions	1.60	0.122
d. Congestion squared and its interactions	1.57	0.128
e. Both Congestion,Congestion squared and Interactions	1.56	0.072
f. Throughput and interaction	208.00	0.000

2. Signs for cargo dummies are relative to ships carrying different cargoes

3. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
120.65	8	1.35

Table 9

Summary of Regression Results For Berths19-21

N=266

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests: WLS** Ratio	P value
Constant	-0.52	-1.29	-0.29	0.75		
Congestion	0.91	2.99	0.64	2.37		
Congestion Squared	-0.81	-2.94	-0.54	-2.19		
Port Wide Congestion	1.52	1.38	1.39	1.35		
Port Wide Congestion Squared	-1.15	-1.41	-0.95	-1.23		
Ln TEUS	0.44	12.02	0.40	9.99		

Notes:

1. F Tests	Ratio	P value
a. Port Wide Congestion & Port Wide Congestion squared	1.10	0.333
b. Congestion & Congestion squared	2.87	0.048

2. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
41.14	1	0.004

Table 10

Summary of Regression Results For Berths 22-23
N=714

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests: WLS** Ratio	P value
Constant	0.52	1.53	0.60	1.90		
Congestion	1.19	3.17	1.34	3.16		
Congestion squared	-1.07	-3.09	-1.19	-3.04		
Port Wide Congestion	4.38	4.25	4.05	4.29		
Port Wide Congestion Squared	-3.24	-4.14	-3.00	-4.16		
Ln Throughput	0.30	7.55	0.28	6.88		
Cargo Dummies					11.83	0.000
Palmoil	-0.56	-4.51	-0.52	-3.92		
Petroleum	-0.01	-0.13	0.03	0.29		
Congestion - Interacted with dummies					2.20	0.111
Palmoil-Congestion	-0.30	-0.50	-0.56	-0.88		
Petroleum-Congestion	-0.73	-1.65	-0.96	-2.05		
Congestion squared - Interacted with dummies					1.78	0.169
Palmoil-Congestion Squared	0.39	0.73	0.64	1.11		
Petroleuml-Congestion Squared	0.62	1.50	0.82	1.89		
Log Throughput - Interacted with dummies					7.39	0.001
Palmoil-Ln Throughput	0.14	2.56	0.14	2.54		
Petroleuml- Ln Throughput	0.16	3.44	0.18	3.85		

Notes:

1. F Tests

	Ratio	P value
a. Port Wide Congestion & Port Wide Congestion squared	9.46	0.000
b. Congestion & Congestion squared	5.01	0.007
c. Congestion and its interactions	5.48	0.001
d. Congestion squared and its interactions	4.83	0.002
e. Both Congestion,Congestion squared and Interactions	3.09	0.005
f. Throughput and interaction	189.09	0.000

2. Signs for cargo dummies are relative to mix cargo

3. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
42.21	1	0.004

Table 11

Summary of Regression Results For Berths 24-25
N=151

Variables	**OLS** Coeff.	t-stats	**WLS** Coeff.	t-stats	**F-tests: WLS** Ratio	P value
Constant	0.90	0.77	0.03	0.02		
Congestion	0.09	0.11	0.52	0.69		
Congestion Squared	0.15	0.18	0.03	0.04		
Port Wide Congestion	9.06	2.53	10.54	2.90		
Port Wide Congestion Squared	-6.86	-2.49	-7.91	-2.82		
Ln Throughput	0.28	5.84	0.33	6.35		
Cargo Dummies					5.31	0.006
Chemical	-1.27	-3.35	-1.11	-3.13		
General	-0.99	-3.30	-0.68	-1.78		
Congestion - Interacted with dummies					0.11	0.892
Chemical-Congestion	1.67	1.18	0.53	0.41		
General-Congestion	1.99	1.52	0.52	0.37		
Congestion squared - Interacted with dummies					0.40	0.672
Chemical-Congestion Squared	-1.60	-1.25	-0.73	-0.64		
General-Congestion Squared	-2.04	-1.75	-0.96	-0.81		
Ln Throughput - Interacted with dummies					28.71	0.000
Chemical-Ln Throughput	0.53	5.34	0.58	6.93		
General-Ln Throughput	0.42	5.16	0.48	5.30		

Notes:

1. F Tests	Ratio	P value
a. Port Wide Congestion & Port Wide Congestion squared	5.92	0.003
b. Congestion & Congestion squared	22.49	0.000
c. Congestion and its interactions	0.77	0.514
d. Congestion squared and its interactions	0.50	0.681
e. Both Congestion, Congestion squared and Interactions	7.90	0.000
f. Throughput and interaction	115.15	0.000

2. Signs for cargo dummies are relative to wheat cargo

3. Bruesch-Pagan Tests

Test Statistics	Degrees of freedom	Critical value at 5% level of significance
48.78	1	0.004

Perhaps the most important result, in terms of achieving the goal of this research is the joint test of significance on the congestion variables. The WLS coefficients on the congestion, congestion squared variables and their interactions for all subsystems are jointly significantly different from zero at the 1% level (except for berths 12-18 which is significant only at the 10% level); thus, we generally can reject the hypothesis that service time is independent of the occupancy/congestion rate. We also notice that berths 8-11 and 24-25 do not have individual congestion and congestion squared coefficients that are significantly different from zero. However, the joint tests are highly significant which continues to support the hypothesis that service time is dependent on the congestion level. Furthermore, a separate regression consisting only the linear term results in a highly significant coefficient on the congestion variable. The coefficients on the congestion variable are 0.25 with a t-statistic of 3.71 for berths 8-11 and 0.54 with a t-statistic of 6.72 for berths 24-25.

We provide comparative plots of log service time against congestion in Figure 1 and the expressions for the partial derivative of log service time with congestion in Table 12. Table 13 gives the average service time for zero and full congestion levels derived from the estimated regression equation. This table is useful since it provides the exact magnitude of increase in service time as congestion rises to the full occupancy level; and this is especially useful since some of the plots in Figure 1 appear to be flat.

For all subsystems service time initially increases and then declines although the rate of increase and the extent of decline differs across subsystems. In the case of berths 1-2, the early increase in service time appears to be completely wiped out by its later

Table 12

Partial Derivative of Log Service Time With Respect To Congestion and Expressions for Log Service Time as a Function of Congestion

Berths	Partial derivative of Log Service Time with respect to Congestion	Log Service Time as a function of Congestion
1-2	0.77-1.53Congestion	2.51+0.767Congestion-0.768Congestion Squared
5-7A	1.71-1.95Congestion	3.13+1.71Congestion-0.97Congestion Squared
8-11	0.31-0.08Congestion	2.21+0.31Congestion-0.04Congestion Squared
12-18	5.52-6.45Congestion	1.25+5.52Congestion-3.23Congestion Squared
19-21	0.64-1.08Congestion	2.37+0.64Congestion-0.54Congestion Squared
22-23	0.73-1.28Congestion	2.14+0.73Congestion-0.64Congestion Squared
24-25	0.84-0.98Congestion	3.93+0.84Congestion-0.49Congestion Squared

Note: The expressions above are calculated holding Port Wide Congestion, Port Wide Congestion Squared and Log Throughput and Dummy variables at their sample means.

Table 13

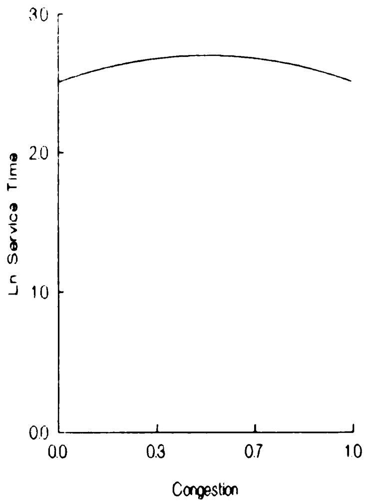
Service time for zero and full congestion.

Sub-System (Berths #)	Service time when there is only one ship at berth (Congestion=0)	Full Occupancy service time (Congestion=1) (% increase in bracket)
1-2	12.3	12.3 (0.0%)
5-7A	22.9	47.9 (109.0%)*
8-11	9.1	11.9 (30.9%)
12-18	5.5	26.3 (378.4%)*
19-21	10.7	11.8 (10.3%)
22-23	8.5	9.3 (9.5%)
24-25	50.9	72.2 (41.8%)

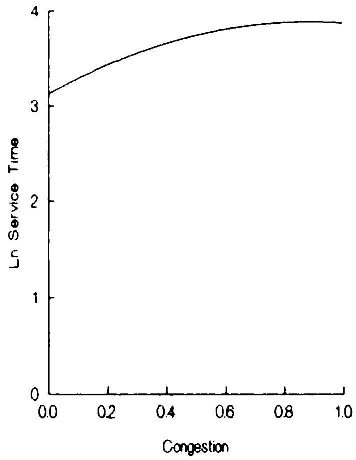
* Please refer to the text below as to why this figure is unrealistic.

Figure 1

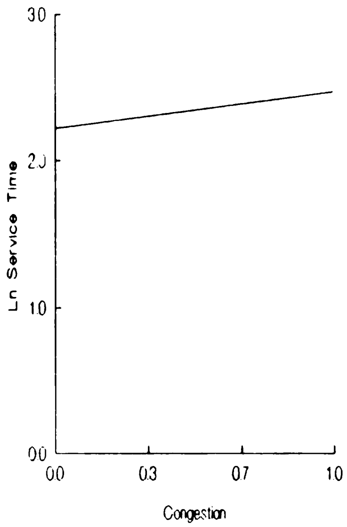
Plots for Log Service Time as a Function of Congestion



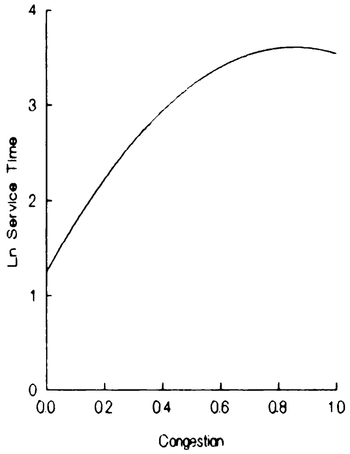
Berths 1-2



Berths 5-7A

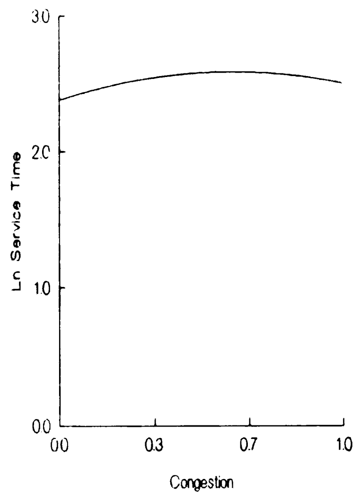


Berths 8-11

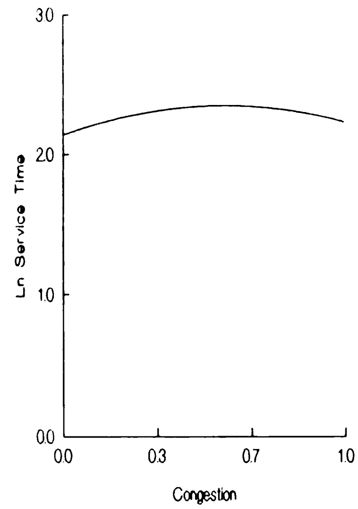


Berths 12-18

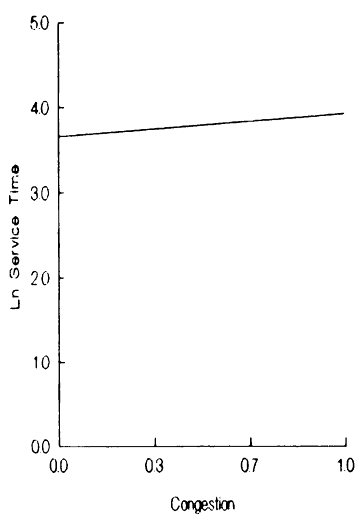
Figure 1 (cont'd)



Berths 19-21



Berths 22-23



Berths 24-25

decline as the congestion rate rises. However, for the other berths the service time never drops far enough to overcome the initial increase like it did for berths 1 and 2.

Two sub-systems appear to give unreasonable estimates as to the increase in service time when there is full occupancy according to Table 13. Berths 5-7A register a leap of 109 percent and, even worse, berths 12-18 register a staggering 378 percent. This strange inference results from a data problem which we alluded to earlier when describing the data. Unlike the other sub-systems, the data from these berths are not well distributed over the range of congestion. In the case of berths 5-7A, there hardly is any data point (less than 5 %) corresponding to congestion levels below 40 percent and for berths 12-18 there is very few below the 45 percent congestion level. So, our inference on the expected service time is potentially erroneous because it is based on projecting the service time downwards using data from higher congestion levels. The projection for zero congestion service time (i.e. only one ship at berth) is more likely to be biased downwards. The reason is as follows: for subsystems involving many berths, like 5-7 and 12-18, it is very likely for the effect to be milder when the congestion index is low than when it is high. If we have data that cover the entire range of congestion, the results would reflect this pattern. However, our data are truncated at the forty or fifty percent level. Therefore, if we project the relationship between service time and congestion downwards, the steeper relationship at higher congestion level will result in under-prediction of the service time when congestion is equal to zero. In other words, the inference that we make for zero congestion service time will be exaggerated downwards. Whilst our results for subsystems involving berths 5-7 and 12-18 are still useful in showing that service time is dependent on congestion rate, we cannot meaningfully use them to predict service times outside the range of congestion in which we have no data. However, we could still make valid inferences from the forty percent level upwards.

Having described the behavior of the congestion variable implied by the regression, we proceed to explain why service time initially increases with congestion and then falls as congestion gets higher. This result at the first instance appears to be counterintuitive. Even if we argue that the port may have increased the inputs for loading/unloading or workers tend to work faster and harder because of greater supervision during busier times, it is apparently not plausible for every ship's service time to fall as congestion increases. Highway congestion is a prime example; travel time increases with congestion, and at a very high level of congestion, the increase can be very dramatic.

This apparent anomaly can be reconciled by taking a second look at the definition of service time. In our analysis, as in any queuing model, service time refers to the duration that a customer spends at a server. In the case of a bank, to take a different example, the service time will be the amount of time a customer spends at a (human) teller. Service time does not include the queuing time. Therefore, the expected total time that a customer spends at his local bank will actually increase during busier period, but his time at a teller can actually fall if the teller works harder when there are more customers to serve. Similarly in the case of port congestion, even though the service time during busy periods may fall (because there are compensating factors at work as the berths get busier), the total waiting time may actually increase due to longer queues.

Apart from understanding the precise definition of service time above, it is also important to realize that a higher congestion rate in our regression analysis does not correspond to a situation where there are more and more ships crowding and exceeding the fixed number of berths. It is not exactly comparable to the case of the highway congestion where more vehicles try to squeeze into a highway with fixed capacity. The congestion rate goes up because in busier periods, every ship tends to spend a relatively longer period of time simultaneously with other ships *at the quay side* for the purpose of loading and unloading.

Now, if indeed the dock workers and stevedores work harder because of greater supervision and/or the port authority provides additional resources when the port is busier, then we would have explained the pattern of service time that we got from the regression. During the less busy periods, relatively speaking, every ship experiences full occupancy less frequently and for a shorter period of time. As the port gets busier, ships spend longer and longer periods of time in full occupancy situation resulting in a higher and higher congestion index. A rising congestion index also corresponds to greater worker supervision and more loading/unloading resources being employed. As the compensating factors increase with the congestion index, there may come a point when their increasing negative impact on service time will cause the service time to start to decline. When this happens, we get a declining pattern of service time over the higher range of the congestion index. Unfortunately, we do not have access to relevant data to test whether workers' productivity and loading/unloading resources increase with the level of congestion.

Having discussed the issue of system congestion, we now move to the port wide congestion variable. The results are mixed since the port wide congestion variable appears to have significant impact on service time for three subsystems only, namely berths 8-11, 22-23 and 24-25. We provide the partial derivatives for log service time with respect to port wide congestion variable and the expression of log service time as a function of port wide congestion in Table 14. Figure 2 gives the plots for the relationship between log service time and port wide congestion.

Berths 8-11 exhibit a rather peculiar U shaped relationship. However, from the lowess plot we notice there are relatively few points below the 40% congestion level. As a result the initial downward sloping portion of the relationship is not really a valid inference given the data that we have. The upward sloping portion can be explained by the resource sharing argument. For berths 22-23 and 24-25, service time initially varies positively with congestion, but later declines. This pattern is similar to the one we found for the subsystem congestion variable. Certainly, we can explain this pattern by invoking the same argument used in explaining the behavior of service time in relation to subsystem congestion. The initial increase in service time is due to a common resource (like pilotage service) being spread more thinly. Later, increases in inputs and supervision of loading/unloading resources, as the port gets more congested, cause service time to decline. In addition, we can offer another explanation for the declining part of the relationship. Recall that pilotage is compulsory at Port Klang. Pilotage refers to the service provided by a pilot to navigate a ship through the approach channel as well as the use of several tug boats to facilitate proper berthing at the dock. Arguably, when there are more ships at the port, there are more tug boats and pilots that are already around in the waterways helping ships to dock. As a result during busier period, any call for service will receive a faster response.

Next we consider the throughput variable. Not surprisingly, service time increases

Table 14

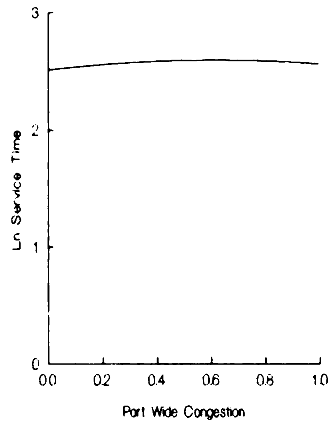
Partial Derivative of Log Service Time With Respect To Port Wide Congestion and Expressions for Log Service Time as a Function of Port Wide Congestion

Berths	Partial derivative of Log Service Time with respect to Port Wide Cogestion (PWC)	Log Service Time as a function of Port Wide Congestion (PWC)
1-2	0.27-0.44PWC	2.51+0.27PWC-0.22PWC Squared
5-7A	1.50-2.68PWC	3.28+1.5PWC-1.34PWC Squared
8-11	-1.27+2.38PWC	2.70-1.27PWC+1.19PWC Squared
12-18	2.01-3.34PWC	1.93+2.01PWC-1.67PWC Squared
19-21	1.39-1.90PWC	2.01+1.39PWC-0.95PWC Squared
22-23	4.05-6.0PWC	1.85+4.05PWC-3.00PWC Squared
24-25	9.32-13.9PWC	1.48+9.32PWC-6.95PWC Squared

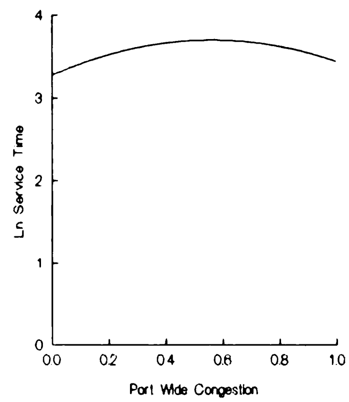
Note: The expressions above are calculated holding Congestion, Congestion Squared and Log Throughput and Dummy variables at their sample means.

Figure 2

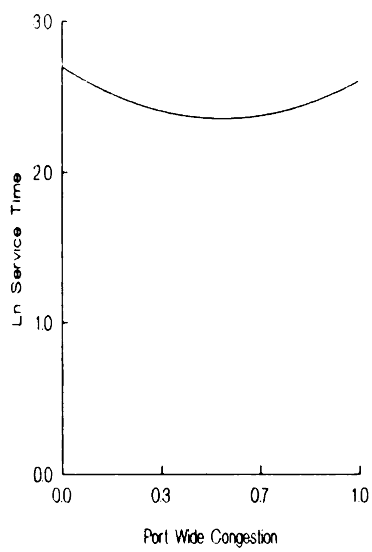
Plots for Log Service Time as a Function of Port Wide Congestion



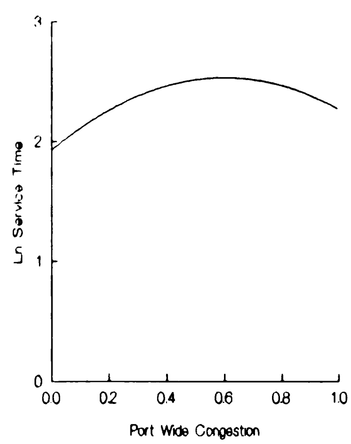
Berths 1-2



Berths 5-7A

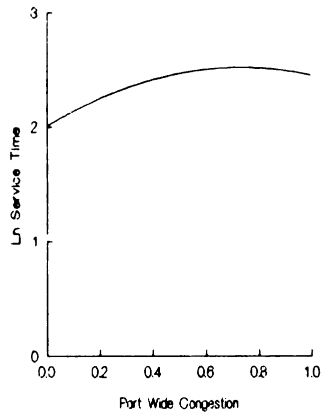


Berths 8-11

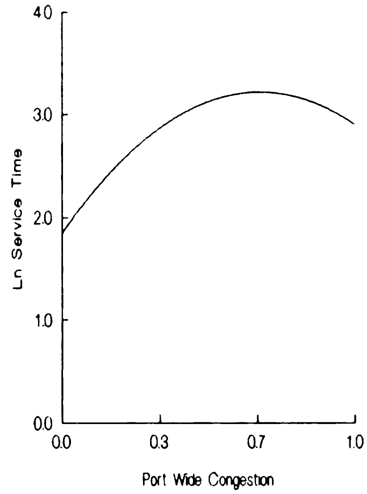


Berths 12-18

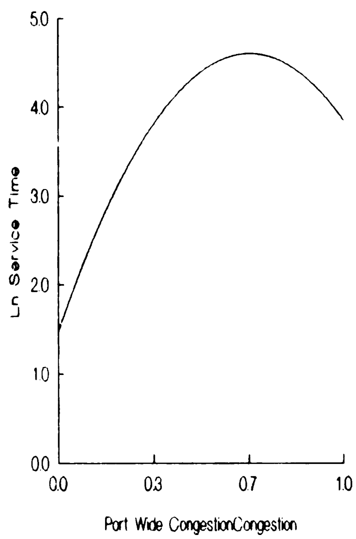
Figure 2 (cont'd)



Berths 19-21



Berths 22-23



Berths 24-25

as throughput rises and the coefficients are consistently significant across all subsystems. The coefficients measure elasticity since we have both service time and throughput in log form. For all berths, they are less than one but greater than zero, indicating that it takes proportionally less increase in time for a given increase in cargo throughput. Table 15 gives us the comparative figures for the percentage increase in service time for a 100% increase in cargo volume.

One possible source of the increasing returns to scale is a common set up time regardless of ships and cargo size: bigger ships may take the same amount of time to be ready for loading/unloading. So, the average service time for loading/unloading tends to drop as the amount of cargo increases. Yet another possibility is that bigger ships are more efficient in handling cargo because they may have better cargo hull design and more loading/unloading inputs of their own to supplement the port side resources.

We also notice that increasing returns to scale are much less pronounced for ships that carry a more heterogeneous cargo as represented by a higher percentage increase in service time for berths 5-7A and 12-18, both of which handle break-bulk cargo. Greater variety of cargo are more difficult to handle thus reducing the advantage of greater volume. It is also possible that bigger break-bulk cargo ships not only carry more of each type of cargo but also bring a greater variety too.

Although the finding of increasing returns to cargo throughput is consistent with economic intuition, there is still some doubt about its magnitude. The elasticity measures between service and throughput hovers around 50%, which looks surprisingly low. The explanation most like lies in the omission of port and ship side loading/unloading inputs

Table 15

Percentage Increase in Expected Service Time for a 100% Increase in Throughput

Sub-System (Berths #)	Increase in Service Time
1-2	40%
5-7A	53%
8-11	48%
12-18	50%
19-21	40%
22-23	40%
24-25	65%

variables. As we argue in the regression model section, both variables tend to bias estimated coefficients on log throughput variable downwards.

Moving on to the cargo dummies, the results show that ships carrying different cargoes do have different expected service times, controlling for other factors. The pattern of variations are also as expected. For example in the case of berths 1-2, a standard cargo like palm oil took a shorter period of time to load/unload relative to mix cargo. The same pattern is also observed for berths 5-7 and 22-23. There is one exception, however. For berths 12-18 cargo type appears not to influence the expected service time. The coefficients on cargo dummies are not significant at the 5% level for both the t and F test.

The coefficients on the interaction variables between congestion and cargo dummies are generally not significant except for berths 1-2. The F test also indicates that the variables are jointly insignificant for those berths. We can conclude, therefore, that in general ships carrying different cargoes do not have their service time impacted differently by congestion. Congestion does affect service time but the impact does not depend on the type of cargo carried.

The results for the interaction between throughput and cargo dummies are always significant and it is consistent across all subsystems except for berths 12-18. Ships carrying different cargoes do have their service time impacted differently by the throughput variables. However, we are unable to identify a general pattern that runs across all subsystems.

Finally, the regression on squared residuals for all subsystems indicate that the OLS violates the homoscedasticity assumption since the statistics for the Bruesch-Pagan test are all significant. Throughput significantly explains variations in the OLS squared residuals for all subsystems. The signs are all negative except for berths 5-7A. This results implies that bigger ships tend to have milder variability in service time.

Conclusion

In this chapter we have shown that the research hypothesis of dependence between service time and occupancy rate cannot be rejected by the data from Port Klang. The magnitude of dependence, however, varies from one subsystem to another. This finding lends evidence against one of the standard assumptions of queuing model as it is applied to port operation.

We also found other factors which systematically affect service time. These factors have, thus far, been ignored in the development of queuing theory as a tool in modelling port operation. Cargo throughput and cargo type are the two most significant. Augmenting the queuing model to handle these factors will certainly improve the prediction and reliability of queuing theory in modelling port operation.

In the next chapter we determine whether variable service time has significant welfare implication through its impact on congestion toll and investment pattern at Port Klang.

Chapter 5

Optimal Congestion Toll and Investment Pattern at Port Klang

In this chapter we calculate the optimal toll charges and investment pattern based on the results of the regression analysis of Chapter 4 and the analytical model of Chapter 3. In the process, we compare the optimal toll charges, timing of investment and the welfare implication arising from the two assumptions on the relationship between service and occupancy rate.

Calculations and Assumptions

There are two service rates used in our calculations; one for less-than-full occupancy and the other for full occupancy. By having only two rates, we assume that the service time changes only when full occupancy is achieved. By making such a simplifying assumption we essentially ignore any increase in service time for less than full occupancy. We have to make this assumption because the analytical model becomes unwieldy if we try to solve a model with more than two service rates. At the same time, we also choose the most conservative estimates for the increase in service time for each subsystem. The service times are listed in Table 10 (Chapter 4) earlier. Berths 5-7 and 12-18 are dropped because of the data problem described in the previous section. We also drop berths 1-2 because according to our regression at full occupancy the expected service time is equal to the zero occupancy level. This leaves us with four subsystems in our calculation involving berths 8-11, 19-21, 22-23 and 24-25.

Estimating the optimal congestion toll requires that we solve the following three equations which were first derived in Chapter 3. Using

$$\frac{Fc\mu}{C_q} = Z_p \tag{1}$$

and

$$Z_p = \frac{\left\{[VS] - U\rho^{-1}\left[(z^* - c)(-R)(1-\rho)^2 + (z^* - c)(-s + \rho S) + \rho S(1 - \rho^{z^* - c})\right]\right\}}{\left\{T\rho\left[R(1-\rho) + S\right]\right\}} \tag{2}$$

we can derive the optimal queue size, k^*. We, in turn, uses k^* to estimate the optimal toll by utilizing,

$$F - C_q\left(\frac{k^* + 1}{c\mu}\right) < t \le F - C_q\left(\frac{k^*}{c\mu}\right) \tag{3}$$

where $R = \sum_{n=0}^{c-1} \frac{r_1^n}{n!}$, $S = \frac{\rho r_1^{c-1}}{(c-1)!}$, $T = \frac{r_1^{c-1}}{(c-1)!}$, $U = \frac{r_1^{c-1}\rho^2}{(c-1)!(1-\rho)^2}$, $V = \sum_{n=0}^{c-1}(c-n)\frac{r_1^n}{n!}$

F is the penalty for loading/unloading at an alternative port.

C_q is the cost of waiting per unit time per ship.

c is the number of berths in a sub-system.

μ is the service rate per ship per unit time when all berths are occupied.

λ is the arrival rate per unit time

$$\rho = \frac{\lambda}{c\mu} \quad \text{and} \quad r_1 = \frac{\lambda}{\mu_1}$$

k^p is the privately optimal maximum queue length.

k^* is the socially optimal allowable queue length.

Equation 1 gives us the optimal queue strategy of private operators. It is equal to the ratio of the expected penalty cost to the expected queue cost per ship. Hence, if the penalty cost of is twice as large as the expected queuing cost, ship operators would only be willing to join the queue if the current queue length is two or fewer. A ship operator would not join a longer queue because he is better off going to an alternative port. Equation 2 provides the optimality condition for determining the socially optimal queue length given the privately optimal queue strategy derived from Z_p. Once Z^* is found, equation 1 is used to arrive at the optimal congestion toll. This condition was derived in Chapter 3. Equation 3 says that the optimal toll, t, must be set such that ship operators will choose the socially optimal queue length (k^*). Choosing otherwise will not be optimal since the expected cost of queuing plus toll, $C_q\left(\frac{k^*+1}{c\mu}\right)+t$, is greater than the penalty cost.

To arrive at the optimal investment pattern we compare the discounted benefits against the discounted costs of adding a new berth. The economic net present value for the marginal investment can be written as:

$$NB_t = \sum_{t=0}^{T} \frac{1}{(1+r)^t} \left\{ (S_{t,n} - S_{t,n+1}) - (K_t + M_t) \right\}$$

where $S_{t,n}$ is the total sum of waiting and penalty cost for period t when there are n berths in service.

K_t is the capital outlay

M_t is the maintenance cost

r is the discount rate

A new berth should be constructed whenever the net present value is positive and when the cost savings made from the investment is greater than the variable cost.

We have all the data required to perform our calculations except for an estimate of the penalty incurred by balking ships. We arbitrarily assume a penalty level of half the waiting cost per period. Later, we perform a sensitivity analysis to check the robustness of the assumption. The calculation requires current data as well as projections on ship calls, waiting time cost, penalty cost, capital cost, maintenance cost and discount rate. Ship calls are expected to grow at a rate of 6% per year based on Port Klang Authority own projection (Abdul Kadir, 1994). Our estimates of the waiting costs for different types of ships are based on the average rental cost of ships per unit time multiplied by the expected waiting time. The rental costs were gathered from several shipping agents in Malaysia. We must mention, however, that the data for rental rates are based on a very general averaging method. Rental rate is a function of ship characteristics and our average does not take these differences into account. For estimate of the maintenance cost which amounts to US$400 000 per year, we used historical data supplied by the statistical department of the port, with a five percent yearly increment to take care of inflation. The discount rate is arbitrarily set at 10%. The capital cost was also taken from estimates given by the same department at US$20 million per new berth. We also assume that at the end of the economic life of the berth, T, it has a salvage value of zero which is a reasonable assumption for a specialized investment. T is estimated to equal 35 years based on the depreciation rate employed by the port authority in preparing its financial statement. The base year for our calculation is 1995. Using these figures and the calculated sum of waiting and penalty costs, we present the results in Table 16. Finally, note that in our calculation we ignore the cost expended to reduce service time during busier periods. This amount should have been included because it is part of the total cost. Unfortunately, the data are not available and our model does not have the flexibility of incorporating this cost. We explain the likely impact of this omission on the estimates below.

For each sub-system we have two separate tables describing the impact of variable service time on optimal tolls and investment pattern during a ten year period. The first table gives us some idea on the difference in optimal toll charges and timing of investment and the second provides us with the estimated magnitude of welfare gain. The beginning year in the first column is the year when an additional investment (i.e. one more from the existing number of berths for each subsystem) was or will be due under a variable service assumption. The second column compares the optimal number of berths over time, under both assumptions of variable and constant service time. The third column provides the optimal allowable queue length where the capital stock is set assuming constant service time. The last three columns list the size of toll charges, again under both assumptions, and the percentage difference between them.

Table 16

Timing of Investments, Congestion Tolls and Queue Size Under Different Assumptions on Service Times-Berths 8-11 (10-year period)

Year	Optimal number of berths for variable service (Constant Service Time in bracket)	Optimal Allowable Queue Length for Variable Service Time (constant Service Time in bracket)	Optimal Toll if Service Time is Variable ($)	Optimal Toll if Service Time is Constant ($)	Percentage Difference in Toll Charges between Variable and Constant Service Time (Average=50.0)
1992	5 (4)	19(30)	5445	3321	64.0
1993	5 (4)	18(28)	6354	3984	59.5
1994	5 (4)	17(27)	7341	4705	56.0
1995	5 (4)	16(26)	8410	6035	39.4
1996	5 (4)	15(25)	9568	6912	38.4
1997	6 (5)	22(34)	8111	5386	50.6
1998	6 (5)	20(32)	9817	6162	59.3
1999	6 (5)	19(31)	10991	7536	45.8
2000	6 (5)	18(30)	12258	8472	44.7
2001	6 (5)	16(28)	14377	10069	42.8

Welfare Impact of Incorrect Toll Charges and Timing of Investment-Berths 8-11 (10-year period)

Year	Sum of Waiting and Penalty Costs ($) Under Optimal Investment and Toll	Sum of Waiting and Penalty Costs ($) Under Incorrect Investment and Toll	Absolute Gain in Welfare ($) -Net of Additional Maintenance	Percentage Gain in Welfare (Average = 11.1)
1992	2672142	3057320	39643	1.5
1993	3011492	3534698	160394	5.3
1994	3404273	4123232	338007	9.9
1995	3863150	4865256	602106	15.6
1996	4038480	5825004	946524	23.4
1997	4543087	5056396	72309	1.6
1998	5126451	5852360	262859	5.1
1999	5808020	6849884	555662	9.6
2000	6615299	8140752	1014940	15.3
2001	7588706	9879394	1754650	23.1

Table 16 (cont'd)

Timing of Investments, Congestion Tolls and Queue Size Under Different Assumptions on Service Times-Berths 19-21 (10-year period)

Year	Optimal number of berths for variable service (Constant Service Time in bracket)	Optimal Allowable Queue Length for Variable Service Time (constant Service Time in bracket)	Optimal Toll if Service Time is Variable ($)	Optimal Toll if Service Time is Constant ($)	Percentage Difference in Toll Charges between Variable and Constant Service Time (Average=10.1)
2000	4(3)	12(15)	10079	8569	17.6
2001	4(3)	12(15)	10583	8998	17.6
2002	4(3)	12(14)	11112	10654	4.3
2003	4(3)	11(13)	13098	12453	5.2
2004	4(4)	17(20)	11312	10582	6.9
2005	4(4)	16(20)	13060	11111	17.5
2006	5(4)	16(19)	13713	12766	7.4
2007	5(4)	15(18)	15703	14559	7.9
2008	5(4)	14(17)	17858	16499	8.2
2009	5(5)	20(24)	16019	14651	9.3

Welfare Impact of Incorrect Toll Charges and Timing of Investment-Berths 19-21 (10-year period)

Year	Sum of Waiting and Penalty Costs ($) Under Optimal Investment and Toll	Sum of Waiting and Penalty Costs ($) Under Incorrect Investment and Toll	Absolute Gain in Welfare ($)-Net of Additional Maintenance	Percentage Gain in Welfare (Average = 4.2%)
2000	3670948	4250552	69092	1.9
2001	4122726	4883755	224991	5.5
2002	4638595	5644571	443136	9.6
2003	5230730	6568966	747254	14.3
2004	5914751	5914777	26	0.0
2005	6711115	6711233	118	0.0
2006	6924051	7647609	39423	0.6
2007	7778112	8761697	265242	3.4
2008	8757541	10106337	594536	6.8
2009	9888980	9888999	19	0.0

Table 16 (cont'd)

Timing of Investments, Congestion Tolls and Queue Size Under Different Assumptions on Service Times-Berths 22-23 (10-year period)

Year	Optimal number of berths for variable service (Constant Service Time in bracket)	Optimal Allowable Queue Length for Variable Service Time (constant Service Time in bracket)	Optimal Toll if Service Time is Variable ($)	Optimal Toll if Service Time is Constant ($)	Percentage Difference in Toll Charges between Variable and Constant Service Time (Average=4.6)
1999	3(2)	11(13)	8834	7705	14.6
2000	3(2)	11(12)	9575	9382	2.1
2001	3(3)	19(21)	7514	7365	2.0
2002	3(3)	18(20)	8928	8683	2.8
2003	3(3)	18(20)	9375	9117	2.8
2004	3(3)	17(19)	10989	10619	3.5
2005	3(3)	17(19)	11538	11150	3.5
2006	3(3)	16(18)	13377	12861	4.0
2007	4(3)	15(17)	16140	14716	9.7
2008	4(3)	15(16)	16947	16724	1.3

Welfare Impact of Incorrect Toll Charges and Timing of Investment-Berths 22-23 (10-year period)

Year	Sum of Waiting and Penalty Costs ($) Under Optimal Investment and Toll	Sum of Waiting and Penalty Costs ($) Under Incorrect Investment and Toll	Absolute Gain in Welfare ($)-Net of Additional Maintenance	Percentage Gain in Welfare (Average = 10.1%)
1999	2241783	3044655	316670	14.1
2000	2500400	3524159	513246	20.5
2001	2790013	2972007	181994	6.5
2002	3114728	3348551	233823	7.5
2003	3479328	3780666	301338	8.7
2004	3889427	4279184	389758	10.0
2005	4351674	4857974	506300	11.6
2006	4874028	5535068	661040	13.6
2007	5466124	6334523	150056	2.7
2008	6139794	7288769	394716	6.4

Table 16 (cont'd)

Timing of Investments, Congestion Tolls and Queue Size Under Different Assumptions on Service Times-Berths 24-25 (10-year period)

Year	Optimal number of berths for variable service (Constant Service Time in bracket)	Optimal Allowable Queue Length for Variable Service Time (constant Service Time in bracket)	Optimal Toll if Service Time is Variable ($)	Optimal Toll if Service Time is Constant ($)	Percentage Difference in Toll Charges between Variable and Constant Service Time (Average=69.3)
1991	3(2)	4(6)	1087	504	115.5
1992	3(2)	4(6)	1141	530	115.4
1993	3(2)	4(6)	1198	556	115.4
1994	3(2)	4(5)	9444	6355	48.6
1995	3(2)	4(5)	9916	6673	48.6
1996	3(3)	5(8)	8908	5946	49.8
1997	4(3)	5(8)	9353	6243	49.8
1998	4(3)	5(8)	9821	6555	49.8
1999	4(3)	5(8)	10312	6883	49.8
2000	4(3)	5(8)	10828	7227	49.8

Welfare Impact of Incorrect Toll Charges and Timing of Investment-Berths 24-25 (10-year period)

Year	Sum of Waiting and Penalty Costs ($) Under Optimal Investment and Toll	Sum of Waiting and Penalty Costs ($) Under Incorrect Investment and Toll	Absolute Gain in Welfare ($)-Net of Additional Maintenance	Percentage Gain in Welfare (Average = 10.9%)
1991	1324546	1806231	152604	11.5
1992	1493576	2066922	227811	15.3
1993	1686744	2367966	318410	18.9
1994	1908056	2624218	335209	17.6
1995	2162269	2997394	435124	20.1
1996	2455035	2467875	12839	0.5
1997	2390089	2812596	18493	0.8
1998	2685370	3213648	65228	2.4
1999	3021609	3681673	173862	5.8
2000	3405956	4229361	312892	9.2

In the second table we measure the percentage gain in welfare, acknowledging that service times are actually variable. In column two, we have the sum of waiting and penalty costs assuming that the toll charges and investment are at their optimal values; and the calculation is done using variable service time. We then have, in column three, the sum of waiting and penalty cost using the toll charges and investment pattern arising from the wrong assumption that the service time is constant. We also provide the percentage difference between the two sums.

Having described Table 14 in detail, we will now try to identify some general findings that can be derived from the table. Notice first, that the toll charges calculated under the assumption of variable service time are always greater than that under constant service time. The reason is fairly evident. With variable service time there will be a greater congestion externality because the decline in berth performance will cause longer waiting either in queue or in service. Higher toll charges are therefore necessary to internalize this greater congestion externality. This observation implies that toll charges will be set too low (such that queue entry is under discouraged) if we fail to recognize variable service time. The differences in optimal toll charges between the two assumptions average around 33.5%; a difference which we cannot easily ignore. Berths 24-25 register the highest average percentage difference of 69.3% and berths 22-23 has the lowest average difference of 4.6%.

We also notice that a bigger wedge between the socially optimal allowable queue lengths always results in a greater percentage difference in the amount of congestion tolls, with the number of berths held constant. For example, in 1992 the difference in allowable queue length for berths 8-11 was 11 and in 1993 it dropped to 10. The corresponding difference in toll charges also fell from 64% to 59.5%. Similarly, when the difference in allowable queue length declines from 3 to 2 for berths 10-21 in 2002, the percentage difference in toll charges also decline from 17.6% to 4.3%. This pattern should be expected because a greater wedge means a bigger difference in the marginal congestion 'recognized' under the two assumptions which in turn translates into a greater difference in toll charges.

From the results we can also deduce that the optimal number of berths under the assumption of variable service time is always greater than (or at least equal to -because of lumpy investment) that under constant service time. This finding illustrates our hypothesis that failure to recognize variability in service time generally results in under investment in berthing facilities. With variable service time, a greater congestion externality is recognized in the calculation of the cost and benefits of a an additional investment. Decline in berth performance leads to greater congestion because ships not only have to wait longer while being serviced but also cause other ships to wait longer for service. Both factors make earlier investment more valuable in the variable service time scenario.

However, we may ask the question why are there still some periods in which both scenarios results in the same number of berths. For example in the year 2009 for berths 19-21, the optimal number of berths is five under both assumptions. Another example is from berths 22-23 where between the year 2001 to 2006 the number of berths is equal to 3. The answer lies in the lumpiness of investment. Since investment can only be made in a discreet and lumpy fashion, there will be times in which we are 'stuck' with a certain number of berths. In other words, only if investment were continuously variable would we have different 'amount' of berths, at any one time, as the rate of arrival increases.

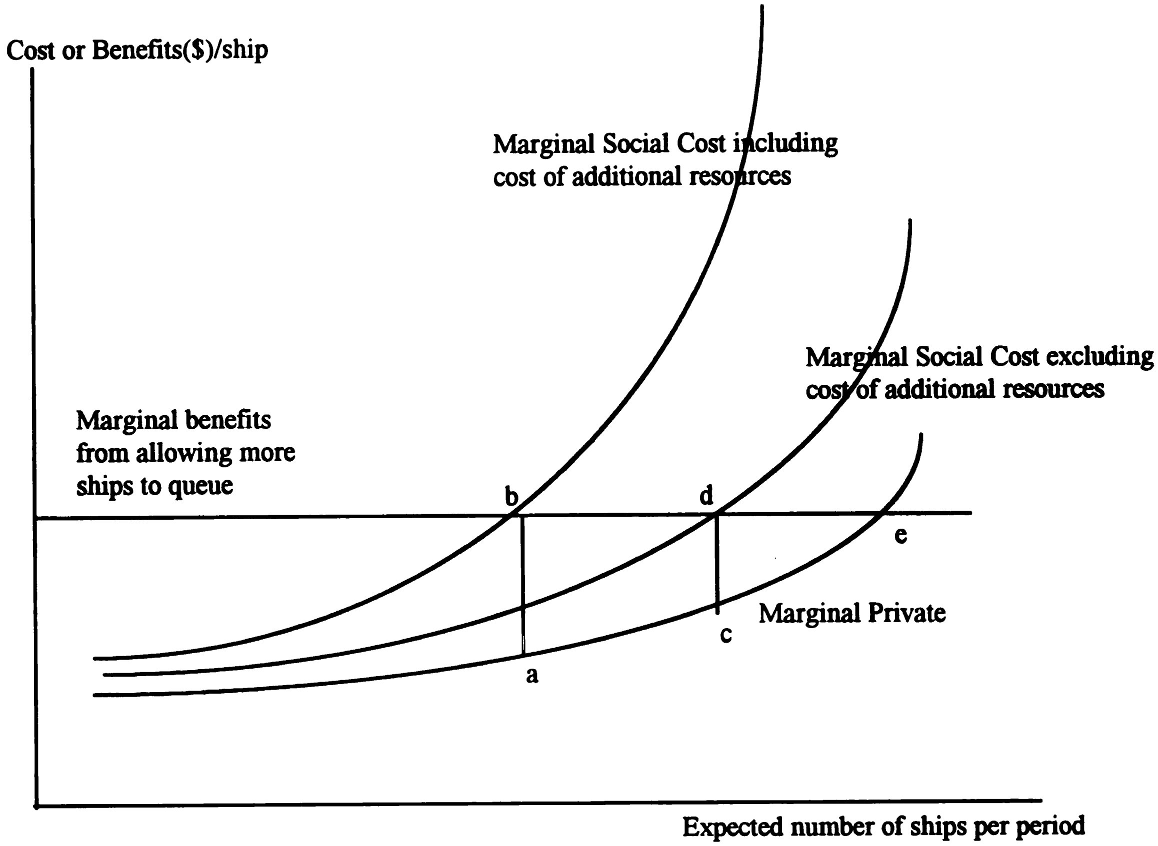
To aid our understanding of the analysis above it is worthwhile to include a graphical exposition of the discussion. Figure 3 provides a graphical representation of the analysis to arrive at the optimal congestion toll. The horizontal line represents the marginal benefit of allowing longer queue per ship. It is equal to the penalty cost saved by allowing an additional ship to queue. The lowest upward sloping curves represent the marginal private cost incurred by individual ship operator. It is upward sloping because the expected waiting time increases as longer queues are allowed to form. The other two upward sloping curves are the marginal social cost curves: one with and the other without the cost of additional resources to speed up loading during busy periods.

We concentrate our discussion first with the optimum point ignoring the cost of additional resources employed by the port operator to speed up the loading/unloading process during busier periods. In this case the marginal social cost is the sum of private cost plus the congestion externality incurred when longer queues are allowed to form. The equilibrium that obtains under no-toll condition will be point 'e'. For social optimum the equilibrium should be at point 'd' with the toll charge equals to the length 'cd'. This is the optimum point that we mathematically derive from our calculations in this chapter.

However, as we indicated earlier the empirical relationship between service and occupancy rate indicates that it is very likely that the port authority does employ increasing amount of resources to speed up loading process as occupancy rises. If this is the case then the marginal cost curve has not embodied the true marginal social cost. As the amount of resources expended is positively related to the occupancy rate, the correct

Figure 3

Optimal Congestion Toll



marginal social cost curve should be above the current marginal cost curve. With this higher marginal cost curve, the congestion toll ought to have been higher (equals to the length 'ab') and the expected ships allowed per period ought to be smaller. Notice also that the higher marginal cost means that investment will become justifiable even earlier because the benefits of a new berth is greater at any point in time. Our calculation of the welfare impact of recognizing variable service time is therefore biased downwards.

Finally, we conclude by highlighting the idea that differences in congestion tolls and investment pattern alone, are not sufficient to justify taking variable service time into consideration in pricing and investment policy decisions. What matters most is the impact these differences have on welfare, which can only be ascertained by comparing the sum of waiting and penalty costs under the two scenarios. The comparison is presented in the tables above. We found the average welfare loss for the four subsystems under consideration is close to ten percent. Berths 8-11 have the highest average gain of 11.1% and berths 19-21 the lowest with 4.5%. However, we must remember the welfare gain from recognizing variable service time is, by design, understated because our calculations are done using the most conservative estimates of the increase in service time when occupancy rate rises. This estimates are conservative because we ignore any increase in service time before full occupancy is achieved. We also ignore the additional cost incurred in reducing service time during busier times at the port. Therefore, the actual welfare implication is certainly larger than our existing estimates.

Before performing the sensitivity analysis, it is also interesting to discuss, at this juncture, the likely implication on the results if the goal of the port operator is to maximize revenue. From Figure 3, ignoring the cost of additional resources, the most that can be charged by the port operator is the vertical distance between the marginal private cost curve and the horizontal penalty cost line. The revenue collected will then be the vertical distance times the expected number of ship arrivals. For illustration, if the charge is equal to the distance ‘cd’, the revenue generated is equal to ‘cd’ times the corresponding expected number of ship arrivals. A revenue maximizing port operator will choose a charge such that the area of the rectangle is maximized. Without further algrbraic analysis we could not easily identify any obvious point on the graph to represent this revenue maximizing point.

Sensitivity Analysis: Varying the Penalty Cost

In calculating the optimal toll and investment pattern we produced many estimates on the relevant variables in the calculation. Many of these estimates are based on information from Klang Port Authority or published literatures. However, the penalty cost of balking ships are chosen arbitrarily without any supporting reference. It is therefore prudent to perform a sensitivity analysis to see if the results still generally hold under different penalty levels. For the purpose of illustration only, we have chosen berths 8-11 as the setting for the analysis.

We present a summary of the results in Table 17, 18, 19 and 20. The current estimate of penalty cost is US$20 000 which is half the average waiting cost of a ship per week. We add three other levels of penalty: US$10000, US$30000 and US$40000.

Table 17

Timing of Investments For Different Penalty Levels-Berths 8-11 (10-year period)

Year	Optimal number of berths for variable service (Constant service time in bracket)	Optimal number of berths for variable service (Constant service time in bracket)	Optimal number of berths for variable service (Constant service time in bracket)	Optimal number of berths for variable service (Constant service time in bracket)
	Penalty=10000	Penalty=20000	Penalty=30000	Penalty=40000
1992	5(4)	5(4)	5(4)	5(4)
1993	5(4)	5(4)	5(4)	5(4)
1994	5(4)	5(4)	5(4)	5(4)
1995	5(4)	5(4)	5(4)	5(4)
1996	5(4)	5(4)	5(4)	5(4)
1997	6(5)	6(5)	6(5)	6(5)
1998	6(5)	6(5)	6(5)	6(5)
1999	6(5)	6(5)	6(5)	6(5)
2000	6(5)	6(5)	6(5)	6(5)
2001	6(5)	6(5)	6(5)	6(5)

Table 18

Allowable Queue Length Under Different Penalty Levels-Berths 8-11 (10-year period)

Year	Optimal allowable queue length for variable service (Constant service time in bracket)	Optimal allowable queue length for variable service (Constant service time in bracket)	Optimal allowable queue length for variable service (Constant service time in bracket)	Optimal allowable queue length for variable service (Constant service time in bracket)
	Penalty=10000	Penalty=20000	Penalty=30000	Penalty=40000
1992	10(15)	19(29)	27(43)	36(49)
1993	10(15)	18(28)	26(41)	33(48)
1994	10(14)	17(27)	24(39)	31(47)
1995	9(14)	16(25)	22(37)	29(46)
1996	9(13)	15(24)	20(35)	26(46)
1997	12(18)	22(34)	31(49)	41(54)
1998	12(18)	20(33)	29(48)	38(54)
1999	11(17)	19(31)	27(46)	35(54)
2000	11(16)	18(30)	25(43)	32(53)
2001	10(16)	16(28)	22(41)	28(53)

Table 19

Toll Charges Under Different Penalty Levels-Berths 8-11 (10-year period)

Year	Optimal toll charges for variable service (Constant service time in bracket)	Optimal toll charges for variable service (Constant service time in bracket)	Optimal toll charges for variable service (Constant service time in bracket)	Optimal toll charges for variable service (Constant service time in bracket)
	Penalty=10000	Penalty=20000	Penalty=30000	Penalty=40000
1992	2267(1306)	5445(3321)	9229(5337)	12407(11137)
1993	2380(1371)	6354(3984)	10328(6597)	14938(12190)
1994	2500(1961)	7341(4705)	12182(7970)	17023(13321)
1995	3327(2059)	8410(6035)	14196(9464)	19279(14535)
1996	3493(2737)	9568(6912)	16381(11087)	22456(15262)
1997	3281(2089)	8111(5386)	13560(9166.5)	18390(17776)
1998	3445(2193)	9817(6162)	15539(10131)	21261(18665)
1999	4300(2836)	10991(7536)	17682(11703)	24373(19598)
2000	4515(3537)	12258(8472)	20001(13966)	27743(21137)
2001	5494(3713)	14377(10069)	23260(15838)	32143(22194)

Table 20

Welfare Impact of Incorrect Toll Charges and Timing of Investment Under Different Penalty Levels-Berths 8-11 (10-year period)

Year	Welfare Gains by Recognizing Variable Service Time	Welfare Gains by Recognizing Variable Service Time	Welfare Gains by Recognizing Variable Service Time	Welfare Gains by Recognizing Variable Service Time
	Penalty=10000	Penalty=20000	Penalty=30000	Penalty=40000
1992	38339	39659	39659	39659
1993	156893	160394	160469	160469
1994	324487	338006	338484	338484
1995	568887	602106	604835	604840
1996	894167	999436	1013369	1013456
1997	70934	72309	72320	72320
1998	258589	262859	262956	262956
1999	537464	555661	556668	556668
2000	944635	1014940	1019997	1020009
2001	1565659	1754649	1779248	1779534

Table 17 compares the optimal number of berths for a ten year period; Table 18 gives a comparison of the optimal allowable queue length; Table 19 provides the differences in toll charges; and Table 20 gives the comparison of absolute welfare gain under different penalty levels. We learn from Table 17 that the penalty level has no impact on the pattern of optimal investment. The timing of investment is exactly the same for all levels of penalty. The optimal number of berth is 5 from 1992 to 1996 and then 6 from 1997 to 2001 for variable service time. The lumpiness of investment explains this pattern. Table 18, however, presents a completely opposite picture. The optimal allowable queue length is significantly affected by the penalty level. For example, the allowable queue length more or less triple when the penalty increases from US$10000 to US$40000. The optimal toll is also sensitive to penalty levels where it more than doubles when we double the penalty level. Finally from Table 20 we notice that the welfare gains are insensitive to different penalty levels. There is very little variation in the absolute welfare gains as we vary the level of penalty.

With the findings presented in Tables 17 to 20, we conclude that there is certainly a need to get a good estimate of the penalty for balking ships since the optimal toll and queue length are significantly affected by the penalty level. Therefore, our earlier estimates of these variables should be accepted with caution. Nevertheless, the estimates of the welfare gains and timing of investment are still reliable since they are generally insensitive to the penalty level.

Conclusion

In this chapter we have examined the impact of variable service time on optimal toll charges and investment pattern. Our calculations are based on the multi-server state dependent service time model presented in Chapter 3 earlier. We have been careful to ensure that our estimates are calculated in the most conservative manner to avoid exaggeration of the impact of service time on the magnitude of toll charges and investment pattern. We were able to show that recognizing variable service time significantly affects the magnitude of congestion tolls and the timing of optimal investments. Failure to incorporate variable service time leads to non-optimal toll schedule and under investment in port facilities. The welfare gains from different toll charges and investment pattern were conservatively estimated at about 10 percent. We can safely conclude that there are gains to be made by recognizing variable service time at port facilities.

Finally, the sensitivity analysis performed shows that the estimated congestion toll is sensitive to the assumption on the level of penalty but not in the case of investment pattern.

Chapter 6

Summary and Suggestions For Future Research

Port study is one field, among many, where queuing models have been fruitfully applied to aid pricing and investment decisions. Early applications of queuing model can be found in the port investment literature. Later the issue of port congestion pricing was discussed using queuing model as a means of analysis. As the application of queuing model continued to grow, concerns were raised about the validity of the model in representing port operation. One such concern is the arbitrary assumption of constant service time even though there are reasons to believe that port facilities do not exhibit this feature. The concern is further justified because the results derived from the model are sensitive to this assumption. Testing the validity of this assumption has been one of the goal of this research.

We tested the assumption of independence between service and occupancy rate and found, in the case of Port Klang, that the service time is *dependent* on the occupancy rate. This finding lends evidence against the standard independence assumption of queuing theory in modelling port operation. The test was done by means of a regression analysis using individual ship data from Port Klang. The data were obtained from the biggest port in Malaysia, Port Klang. They cover a period of twenty five months (September 1993 to September 1995) and include individual ship data on the duration of stay at the dock for loading/unloading, the quantity of cargo and the types of ship and quantity of cargo.

We classified the berths into seven subsystems serving different ships and cargo. Service time was regressed against several variables deemed to be important like throughput, a measure of berth congestion and port congestion, cargo dummies and several interaction variables. The OLS was heteroscedastic; so we ran weighted least squares regressions instead. In all subsystems, port occupancy were found to systematically influence service time. The relationship found has another interesting feature which is not common in congestion studies: service time initially increases and then decreases as occupancy rate rises. The port operators appear to take compensating action to reduce service time during busy periods. The reduction in service time, however, does not go to the extent to cause the average service time to fall below the single occupancy service time (lowest possible occupancy rate). This observation must be cautiously interpreted, however, since the reduction is in the service time and not the queuing time. It must be differentiated from the observation in the highway congestion, for example, where congestion always results in longer driving time and in many cases the increase is very substantial. We reasoned that this service time-reducing effect can come from two sources. It is plausible to argue that the port operator increases loading/unloading inputs and workers supervision during busy periods. Confirming this hypothesis certainly requires further research on the loading/unloading operation of the port facilities.

Through the same regression analysis we also determined other factors that affect service time which have so far been ignored by the standard queuing model. In the weighted least square regressions performed, the coefficients on the cargo dummies and its some the interaction variables are significant. At the same time the throughput variable is also significant for all subsystem. We conclude therefore that cargo type and quantity of goods loaded/unloaded, to be important determinants of service time, besides port occupancy. Just as in the case of congestion variable, the magnitude of impact varies from one subsystem to another.

There is evidence, though inconclusive, that port wide occupancy level affects service time. Port wide congestion affects service time in only three out of a total of seven subsystems. The results is interesting for subsystems with significant coefficient because it implies that ships congests each other within and without a subsystem.

We also showed that augmenting the queuing model with variable service time can have significant impact on the determination of optimal congestion toll and investment pattern at Port Klang. We used the most conservative method in making the calculations to avoid exaggeration of results especially when we are trying to show that relaxing the assumption of constant service time has significant implications for the determination of the optimal toll and investment pattern. Our conservative calculations indicate that the optimal toll charges under variable service time differs from constant service time anywhere between 1.5 % to 38%. At the same time, failure to recognize variable service time results in under-investment in port facilities. The calculations were done based on the assumption that there is already optimal toll charges imposed. To show that the assumption of constant service time results in general under-investment, optimal investment timings were first calculated under both assumptions. A comparison was then made to determine if there is a general difference in the pattern of investment. This finding conforms with our economic intuition because in the face of growing demand for port services, the congestion costs are always underestimated using constant service time if in actuality service time is increasing with occupancy. Since a marginal investment is only justified when the benefits exceed costs, an additional facility under constant service time only becomes economically viable at a misleadingly later date.

Another major goal achieved in this research is the analytical model that we develop to combine both variable service time and multiple server features in one model, which is an extension of several earlier standard models. It accords neatly with the finding that service time does vary with occupancy rate. In doing so we improve the validity of queuing theory as a means of estimating congestion at transportation facilities with variable service time. The queuing theory is used to derive an expression for queue length under several assumptions to suit the problem at hand. We used this expression later as an input to determine the socially optimal allowable queue length and hence the optimal toll. The object of analysis is to minimize the costs of transporting goods to their destination by choosing a level of allowable queue length. We also allow ship operators to balk to make the model dynamic. With these and other assumptions in the model, we demonstrated that the socially optimal allowable queue length is almost always smaller, and certainly never greater, than the privately optimal queue strategy. This result is consistent with our economic intuition since the pursuit of private interest tends to over-congest a facility through ignoring the negative externality agents impose on one another. Since there is a tendency for private agents to join a queue beyond a length that is optimal, a congestion toll is required to achieve the social optimum.

Using the same analytical model we derive the optimal congestion toll and a rule for optimal investment. The derivation of the optimal congestion toll is a fairly standard procedure. We derive several expressions necessary to find the level of toll charge to compel private operators to internalize the externality they impose on other port users. The charge is chosen such that it is sufficient to discourage entry beyond the optimal level. The investment pattern is determine using the net present value approach. The benefits of additional investment in the form of shorter queuing time and lower penalty costs are compared against the capital and maintenance costs of operating a new berth. The expressions that we derive in this section are eventually used to empirically estimate the optimal congestion toll and investment timings.

Recommendations for Future Research

We propose three areas of investigation as a means to improve the current research as well as a way of exploring new avenues in research on port congestion. Firstly, there is a need to make the queuing model more amenable to heterogeneous ship and cargo characteristics. The ability to handle heterogeneity in ship and cargo characteristics will make the queuing model to predict more accurately the impact of individual ships on congestion cost, which in turn, results in better estimates of the optimal congestion toll and timing of investments. It is not that the current queuing model never allows for variability in service time, but the issue is whether that variability is random or can be explained by some factors in a systematic manner. We have shown that service time can be explained systematically and there is a need to augment the model by incorporating this finding. This will not be an easy undertaking since it effectively requires the model to treat each individual ship as having its own unique service time distribution. A more feasible approach, is to group ships into several classes based on certain vessel and cargo characteristics.

Also related to the issue of ship and cargo heterogeneity is the need to account for differences in waiting and penalty cost. There is a subtle difference between heterogeneity in ship or cargo physical characteristics discussed above and differences that arise from heterogeneous waiting and penalty costs. The queuing model should account for differences due to cost reason so that the resulting toll charges or allowable queue length will be optimal. For example, we must be able to treat, in the model, a big container ship carrying 200 containers for Port Klang differently from a smaller vessel carrying the same number of containers since we would expect the opportunity cost of waiting for the bigger ship is probably greater. Recall that the private and social queue strategies are also a function of waiting and penalty costs. Therefore, charging two ships a same amount of toll may build the wrong incentive for ships to queue. Charging the same amount of toll is unlikely to induce an optimal queuing behavior for ships that differ in their waiting costs.

A second area for future research should investigate whether there is any variability in the productivity and average quantity of loading/unloading resources for different levels of congestion in port. This would test whether our hypothesis that variability in service time is partly due to an association between the levels of productivity or quantity of loading/unloading resources, on the one hand, and congestion levels, on the other hand. We can, for example, test whether there is a significant relationship between throughput per gang hour and the degree of congestion in port. We can also test if there is a more than expected loading/unloading inputs being utilized for periods of higher congestion.

Also related to this issue of loading/unloading resource is the need to estimate the impact of additional loading/unloading resources on service time. Quantifying this relationship is important because if indeed the port operators do try to reduce service time during busy periods, we need to get a handle on how much resources are expended so that the congestion toll will reflect this cost.

Finally, there is also a need to tackle the issue of financing for new capital . In our research we implicitly assume that there will always be sufficient funding available whenever there is a need for a new berth. In addition, we have not shown that the congestion tolls collected are sufficient to pay for the construction of a new project. Sufficient for us to say that an economically viable project may not be financially feasible. During the times when many public ports face budget cuts and some being privatized (as in the case of Port Klang), the issue of financing can be an added constraint worthy of study in formulating policies for optimal investment and congestion toll.

To conclude: we have uncovered a potential limit to queuing models as tools for modelling port operation and estimating congestion costs. In light of the results above, there is still some work to be done to handle heterogeneity of ship and cargo characteristics. It may be the case that we have to try other methods of modelling port operation to better deal with this heterogeneity. While a model should not be expected to capture the entire complexity of real life (because a model is by definition an abstraction of reality), it is misleading to ignore some factors which are potentially crucial in determining private queuing behavior and optimal social queue strategy.

Appendix 1

Appendix 1

Section A

The arrivals and departures of ships can be represented by a birth-death process. When a ship arrives we say a birth occurs and when a ship leaves port after completing service, a death occurs. A birth-death process is Markovian, but it allows for changes of system state at any point in time to be either +1 or -1. With Poisson and negative exponential service time, we can state the followings:

1. Pr(an arrival occurs in an infinitesimal interval of length Δt) = $\lambda\Delta t + o(\Delta t)$[1]
2. Pr(more than one arrival in Δt) = $o(\Delta t)$

For service completion, we replace λ with μ in equations 1 and 2 above.

Now, consider a situation when the system is in state E_n at time $t + \Delta t$ where n is the number of ships in the system. For this to happen the system must have been previously in one of three situations. First, it may be in state E_n in period t and no arrival or departure occurs. Second, it may have been in state E_{n+1} and a departure happens, or thirdly it may have been in state E_{n-1} and an arrival occurs. Therefore, we can write the probability that the system is in state n in period $t + \Delta t$ as:

$$p_n(t+\Delta t) = p_n(t)(1-\lambda_n\Delta t)(1-\mu_n\Delta t)$$
$$+ p_{n+1}(t)(\mu_{n+1}\Delta t)(1-\lambda_{n+1}\Delta t)$$

[1] $o(\Delta t)$ is defined in the following manner. A function $f(\cdot)$ is said to be $o(\Delta t)$ if: $\lim_{\Delta t \to 0} f(\Delta t)/\Delta t = 0$.
What the definition essentially says is that the function $f(\cdot)$ has to go faster to zero than Δt does. This definition guarantees that for any instant in time there can only be one arrival or departure.

$$+ p_{n-1}(t)(\lambda_{n-1}\Delta t)(1-\mu_{n-1}\Delta t)$$

$+ o(\Delta t)$ (see footnote [2] below)

The above equation is only correct for n ≥ 1 because the third term won't apply when n=0.

So, for n=0 the appropriate equation is:

$$p_0(t+\Delta t) = p_0(t)(1-\lambda_0 \Delta t)$$
$$+ p_1(t)(\mu_1\ \Delta t)(1-\lambda_{n+1}\Delta t)$$
$$+ o(\Delta t)$$

Both equations can be rewritten as follows:

$$p_n(t+\Delta t) - p_n(t) = -(\lambda_n + \mu_n)\Delta t p_n(t) + p_{n+1}(t)\Delta t \mu_{n+1} + p_{n-1}(t)\Delta t \lambda_{n-1} + o(\Delta t) \quad \textit{for } n > 1$$

and $$p_0(t+\Delta t) - p_0(t) = -\lambda_0 \Delta t p_0(t) + \mu_1 \Delta t p_1(t) + o(\Delta t) \quad \textit{for } n = 1$$

Dividing both sides of the equations by Δt and taking the limit as Δt→ 0, we will get the differential-difference equations:

$$\frac{dp_n(t)}{dt} = -(\lambda_n - \mu_n)p_n(t) + \mu_{n+1}p_{n+1}(t) + \lambda_{n-1}p_{n-1}(t)$$

and $$\frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t)$$

The stationary or steady-state probability is found by equating both $dp_n(t)/dt$ and $dp_0(t)/dt$ to zero.

$$0 = -(\lambda_n - \mu_n)p_n + \mu_{n+1}p_{n+1} + \lambda_{n-1}p_{n-1}$$

and $$0 = -\lambda_0 p_0 + \mu_1 p_1$$

Or $$p_{n+1} = \frac{(\lambda_n - \mu_n)p_n}{\mu_{n+1}} - \frac{\lambda_{n-1}p_{n-1}}{\mu_{n+1}}$$

---

[2] There should really be more than one term involving $o(\Delta t)$. However, since the sum of several $o(\Delta t)$ is also $o(\Delta t)$, we omit stating all the terms.

and $$p_1 = \frac{\lambda_0 p_0}{\mu_1}$$

The steady-state probability is the probability of finding the system in state n at an arbitrary point in time after the process has reached statistical equilibrium. Once the system has reached statistical equilibrium, the probability of having state n will be independent of time.

Substituting in the equation for n = 1,2,3...... and solving the equations by iteration, it can be shown that the state probability can be expressed as:

$$p_n = \frac{(\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0)p_0}{\mu_n\mu_{n-1}\cdots\mu_1} \qquad \text{for } n \geq 1 \qquad (1)$$

Section B

In this section derive the state probabilities p_n with two additional assumptions beyond those in Section A above. We assume that there are c servers and the performance of the berths declines when all of them are occupied.

Substituting μ_n into (1) above, and knowing that the arrival rate is $\lambda_i = \lambda$ for all i = 1, 2...n; gives us:

$$\begin{aligned} p_n &= \frac{\lambda^n p_0}{n\mu_1(n-1)\mu_1(n-2)\mu_1(n-3)\mu_1\ldots\mu_1} \\ &= \frac{\lambda^n p_0}{n!\mu_1^n} \qquad \text{for } c > n \geq 0 \qquad (2) \end{aligned}$$

where c is the number for berths in the subsystem.

and $$p_n = \frac{\lambda^n p_0}{\mu_n \mu_{n-1} \cdots \mu_1}$$

$$= \frac{\lambda^n p_0}{\left[(c-1)\mu_1 (c-2)\mu_1 \ldots \mu_1\right]\left[c\mu\right]^{n-c+1}}$$

$$= \frac{\lambda^n p_0}{\left[(c-1)!\,\mu_1^{\,c-1} c^{n-c+1} \mu^{n-c+1}\right]} \qquad \text{for } c \le n \qquad (3)$$

We evaluate p_0 by using the fact that the probability densities must sum to one.

$$p_0 \left[\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\,\mu_1^n} + \sum_{n=c}^{\infty} \frac{\lambda^n}{(c-1)!\,\mu_1^{c-1} c^{n-c+1} \mu^{n-c+1}} \right] = 1$$

To reduce clutter, we employ the following definitions and substitute them in the above equations:

$$r = \frac{\lambda}{\mu}, \qquad \rho = \frac{r}{c} = \frac{\lambda}{c\mu}$$

$$r_1 = \frac{\lambda}{\mu_1}$$

So, we get:

$$p_0 \left[\sum_{n=0}^{c-1} \frac{r_1^{\,n}}{n!} + \sum_{n=c}^{\infty} \frac{r_1^{\,c-1} r^{n-c+1}}{(c-1)!\,c^{n-c+1}} \right] = 1$$

We now need to solve for the infinite series $\left[\sum_{n=c}^{\infty} \frac{r_1^{\,c-1} r^{n-c+1}}{(c-1)!\,c^{n-c+1}} \right]$. We can write:

$$\left[\sum_{n=c}^{\infty} \frac{r_1^{\,c-1} r^{n-c+1}}{(c-1)!\,c^{n-c+1}} \right] = \frac{r_1^{\,c-1}}{(c-1)!} \sum_{n=c}^{\infty} \frac{r^{n-c+11}}{c^{n-c+1}}$$

$$= \frac{r_1^{\,c-1}}{(c-1)!} \sum_{m=0}^{\infty} \left(\frac{r}{c}\right)^m \quad \textit{where } m = n\text{-}c\text{+}1$$

$$= \frac{r_1^{\,c-1}}{(c-1)!}\frac{1}{(1-\rho)}$$

Therefore,

$$p_0 = \left[\sum_{n=0}^{c-1}\frac{r_1^{\,n}}{n!} + \frac{r_1^{\,c-1}}{(c-1)!(1-\rho)}\right]^{-1}$$

Section C

In this section we find the expression for p_0 under the modified assumption that we have a finite waiting space. The major difference now is that the probabilities are summed over k instead of ∞. The sum of the state probabilities is still equal to one; hence,

$$p_0\left[\sum_{n=0}^{c-1}\frac{\lambda}{n!\,\mu}^{n} + \sum_{n=c}^{k}\frac{\lambda^n}{(c-1)!\,\mu_1^{c-1}c^{n-c+1}\mu^{\,n-c+1}}\right] = 1$$

Substituting, $r = \dfrac{\lambda}{\mu}$, $\qquad \rho = \dfrac{r}{c} = \dfrac{\lambda}{c\mu}$

and $\qquad r_1 = \dfrac{\lambda}{\mu_1}$,

we get $\qquad p_0\left[\sum_{n=0}^{c-1}\frac{r_1^{\,n}}{n!} + \sum_{n=c}^{k}\frac{r_1^{\,c-1}r^{\,n-c+1}}{(c-1)!c^{\,n-c+1}}\right] = 1$

$$\Rightarrow p_0\left[\sum_{n=0}^{c-1}\frac{r_1^{\,n}}{n!} + \frac{r_1^{\,c-1}}{(c-1)!}\sum_{n=c}^{k}\frac{r^{\,n-c+1}}{c^{\,n-c+1}}\right] = 1$$

$$\Rightarrow p_0\left[\sum_{n=0}^{c-1}\frac{r_1^{\,n}}{n!} + \frac{r_1^{\,c-1}}{(c-1)!}\sum_{m=1}^{k-c+1}\rho^m\right] = 1 \qquad \text{where m=n-c+1}$$

$$\Rightarrow p_0\left[\sum_{n=0}^{c-1}\frac{r_1^{\,n}}{n!} + \frac{r_1^{\,c-1}}{(c-1)!}\left(\sum_{m=0}^{k-c+1}\rho^m - 1\right)\right] = 1$$

$$=> p_0\left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!}+\frac{\rho r_1^{c-1}}{(c-1)!}\frac{(1-\rho^{k-c+1})}{1-\rho}\right]=1 \qquad \text{since } \sum_{m=0}^{k-c+1}\rho^m=\frac{1-\rho^{k-c+2}}{1-\rho}$$

Therefore, the probability of having no ship in the system, p_0, can be expressed as:

$$p_0=\left[\sum_{n=0}^{c-1}\frac{r_1^n}{n!}+\frac{\rho r_1^{c-1}}{(c-1)!}\frac{(1-\rho^{k-c+1})}{1-\rho}\right]^{-1}$$

Appendix 2

Appendix 2

Section A

Our goal is to show that the following inequality is true.

$$\frac{\partial Z_p}{\partial Z^*} = \frac{-U\left[(\rho S - R(1-\rho)^2 - S - \rho^{Z^*-c+1}\rho S \log\rho\right]}{\left\{T\rho^2\left[R(1-\rho)+S\right]\right\}} \geq 0$$

Rewriting the expression into a particular form, we get:

$$\frac{\partial Z_p}{\partial Z^*} = \frac{U\left[(R(1-\rho)^2 + S(1-\rho+\log\rho^{\rho^{Z^*-c+2}})\right]}{\left\{T\rho^2\left[R(1-\rho)+S\right]\right\}} \geq 0$$

First, we evaluate the denominator. Expanding the term $\left[R(1-\rho)+S\right]$ we get:

$$\sum_{n=0}^{c-1}\frac{r_1^n}{n!}(1-\rho)+\frac{\rho r_1^{c-1}}{(c-1)!}$$

$$=\sum_{n=0}^{c-1}\frac{(ac\rho)^n}{n!}(1-\rho)+\frac{(ac)^{c-1}}{(c-1)!}\rho^c$$

$$=-\rho\left[1+ac\rho+\frac{(ac\rho)^2}{2}\ldots\ldots\frac{(ac\rho)^{c-1}}{(c-1)!}\right]+\left[1+ac\rho+\frac{(ac\rho)^2}{2}\ldots\ldots\frac{(ac\rho)^{c-1}}{(c-1)!}\right]$$

$$+\frac{(ac)^{c-1}\rho^c}{(c-1)!}$$

$$=-\left[\rho+ac\rho^2+\frac{(ac)^2\rho^3}{2}\ldots\ldots\frac{(ac)^{c-1}\rho^c}{(c-1)!}\right]+\left[1+ac\rho+\frac{(ac\rho)^2}{2}\ldots\ldots\frac{(ac\rho)^{c-1}}{(c-1)!}\right]$$

$$+\frac{(ac)^{c-1}\rho^c}{(c-1)!}$$

$$=-\left[\rho+ac\rho^2+\frac{(ac)^2\rho^3}{2}\ldots\ldots\frac{(ac)^{c-2}\rho^{c-1}}{(c-2)!}\right]+\left[1+ac\rho+\frac{(ac\rho)^2}{2}\ldots\ldots\frac{(ac\rho)^{c-1}}{(c-1)!}\right]$$

$$=1+\rho(ac-1)+ac\rho^2\left(\frac{(ac)}{2}-1\right)\ldots\ldots\frac{(ac)^{c-2}\rho^{c-1}}{(c-2)!}\left(\frac{(ac)}{(c-1)}-1\right)$$

$$=1+\sum_{c=2}^{C}\frac{(ac)^{c-2}\rho^{c-1}}{(c-2)!}\left(\frac{(ac)}{(c-1)}-1\right)$$

The last expression is definitely greater than zero if we place the condition that $a > \frac{c-1}{c}$.
It turns out that this condition is necessary to make the model sound from an economic point of view, since it is necessary to economically utilize 'c' number of berths instead of 'c-1' berths. Otherwise, it is not economically worthwhile to increase the number of berths to 'c' because introducing an additional berth will actually decrease the overall capacity of the subsystem. For example, in a 3-berth subsystem, we must have a $> \frac{3-1}{3} = 2/3$ since if this is not the case it is better to stick to utilizing only 2 berths instead of 3.

Now, consider the numerator. All of the terms are obviously positive except for the term in the second bracket $1-\rho+\log\rho^{\rho^{z^*-c+1}}$. We now prove that this term is also positive. Let z*=c, since this assumption will result in the smallest possible value for the log variable. The proof is by contradiction:

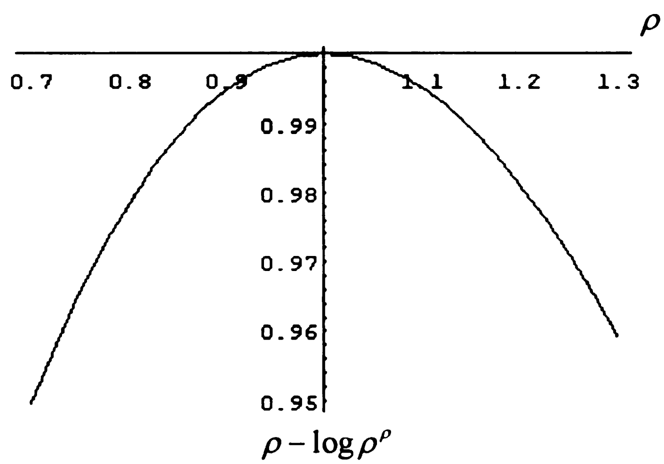
Suppose: $$1-\rho+\log\rho^{\rho} < 0$$
$$\rho-\log\rho^{\rho} > 1$$

But $\rho-\log\rho^{\rho}$ can never exceed 1 and for illustration we demonstrate this fact by the plot below in Figure 4. Therefore, $1-\rho+\log\rho^{\rho^{z^*-c+1}}$ has to be at least zero, for all values of ρ greater than zero. This statement completes the proof that

$$\frac{\partial Z_p}{\partial Z^*} = \frac{U\left[(R(1-\rho)^2 + S(1-\rho+\log\rho^{\rho^{z^*-c+2}})\right]}{\left\{T\rho^2\left[R(1-\rho)+S\right]\right\}} \geq 0$$

FIGURE 4

A plot demonstrating that $\rho - \log\rho^{\rho}$ can never exceed 1



Section B

In this section we provide proof that $Z_p \geq Z^*$. Hence, we want to show that:

$$Z_p - Z^* = \frac{\left\{[VS] - U\rho^{-1}\left[(Z^* - c)(-R)(1-\rho)^2 + (Z^* - c)(-s + \rho S) + \rho S(1 - \rho^{z^* - c})\right]\right\}}{\left\{T\rho\left[R(1-\rho) + S\right]\right\}} - Z^* \geq 0$$

Our effort to manipulate the equation directly from the one above has been unfruitful. Instead try an indirect way in order to show that $Z_p \geq Z^*$. First, we differentiate the above equation and show that the difference between Z_p and Z^* is increasing as Z^* rises. Having done so, we evaluate the function Z_p at the lowest point in the domain, that is when $Z^* = 2$. Recall that in a multi-server queuing model, the lowest value for the number of servers is two. If we can show that Z_p is greater than two than the proof will be completed.

Differentiating Z_p with respect to Z_p, we get:

$$\frac{\partial Z_p}{\partial Z^*} = \frac{U\left[R(1-\rho)^2 + S(1 - \rho + \log \rho^{\rho^{z^* - c + 1}})\right]}{\left\{T\rho^2\left[R(1-\rho) + S\right]\right\}} - 1$$

We have already proven that the numerator is positive. Not only that the term $1 - \rho + \log \rho^{\rho^{z^* - c + 2}}$ has minimum value of zero. Since we are trying to show that the derivative is greater than zero there will be no harm if we choose the minimum value for this term in the numerator. This leaves us to evaluate:

$$\frac{\partial Z_p}{\partial Z^*} = \frac{U\left[R(1-\rho)^2\right]}{\left\{T\rho^2\left[R(1-\rho) + S\right]\right\}} - 1$$

$$= \frac{\frac{T\rho^2}{(1-\rho)^2}\left[R(1-\rho)^2\right]}{\left\{T\rho^2\left[R(1-\rho)+S\right]\right\}} - 1$$

$$= \frac{R}{R(1-\rho)+\rho T} - 1$$

$$= \frac{R}{R+\rho(T-R)} - 1$$

Now, R=$\sum_{n=0}^{c-1} \frac{r_1^n}{n!}$ and T=$\frac{r_1^{c-1}}{(c-1)!}$. So, T<R, therefore, $\frac{R}{R+\rho(T-R)} > 1$, which in turn means that the derivative is greater than zero. So, the difference between Z_p and Z^* is increasing as Z^* rises.

We need only now to show that at the lowest point in the domain, that is when Z^* is equal to 2, corresponds to $Z_p > Z^*$. Evaluating Z_p when $Z^* = 2$, we get:

$$Z_p = \frac{(2+ac\rho)ac\rho^2}{ac\rho^2((1+ac\rho)(1-\rho)+ac\rho^2)}$$

$$= \frac{(2+ac\rho)}{(1+ac\rho)(1-\rho)+ac\rho^2}$$

We now prove by contradiction that Z_p>2. Suppose, $Z_p \le 2$, then:

$$\frac{(2+ac\rho)}{(1+ac\rho)(1-\rho)+ac\rho^2} \le 2$$

$$2 + ac\rho \le 2 - 2\rho + 2ac\rho$$

$$ac \ge 2$$

Recall, however, that ac<2, thus a contradiction. Therefore, Z_p is greater than two, which completes our proof.

Appendix 3

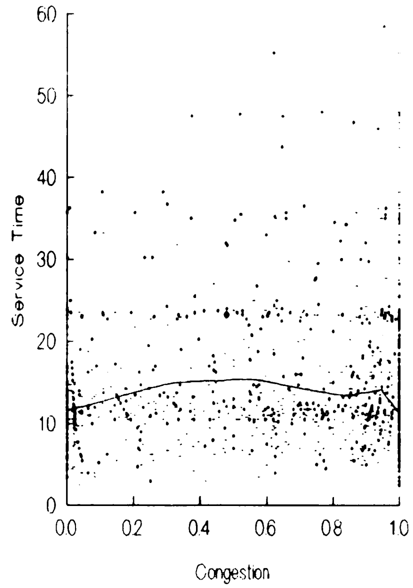
Appendix 3

In this section we present bivariate lowess plots between service time three other variables for all subsystems. Since the plots are bivariate, they are only used to get a general idea on the probable relationship between the two variables. Figure 5 display the relationship between service time and the congestion variables. Notice that the relationship is probably non linear and in particular, service time initially increases and then decline which suggest the inclusion of a quadratic term in the regression. There also appears to be a rapid decline in service time at high congestion levels except for berths 24-25. Due to the difference in scale between the x and y axes, the relationships look flatter than they should.

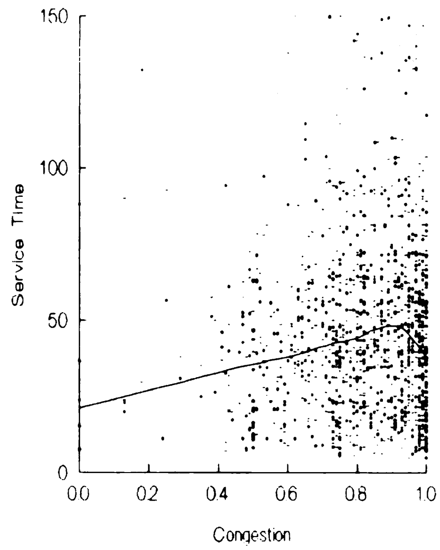
In Figure 6 we have the plots for service time against port wide congestion variables. Note that the relationship appears to be quadratic. Notice also the lack of data for congestion level below the 30% level.

Finally in Figure 7 display the relationship between service time and throughput variables. The is some evidence of increasing returns to scale since the average service time appears to decline as throughput increases. This pattern is consistently manifested by all subsystems. This suggests that we use a functional form which allows for decreasing returns to scale for throughput.

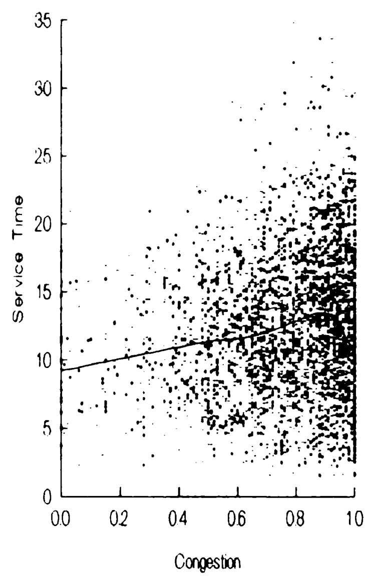
Figure 5
Lowess Plots: Service Time vs Congestion Variables for All Berths



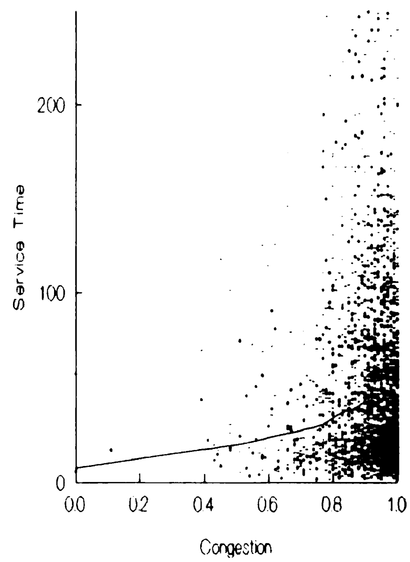
Berths 1-2: Service vs Congestion



Berths 5-7A: Service vs Congestion

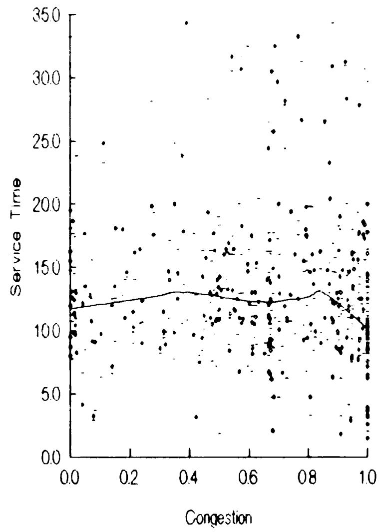


Berths 8-11: Service vs Congestion

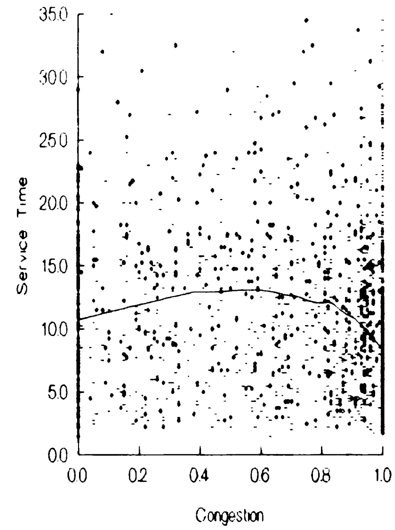


Berths 12-18: Service vs Congestion

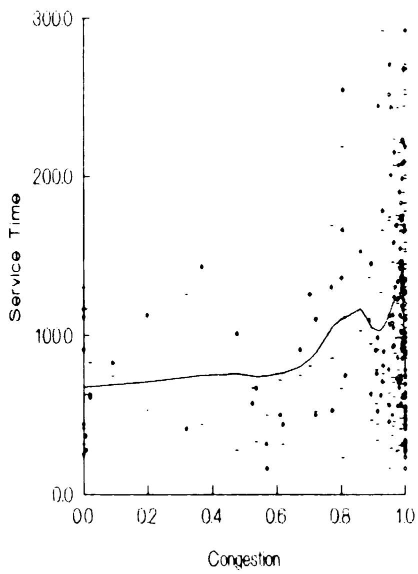
Figure 5 (Cont'd)



Berths 19-21: Service vs Congestion

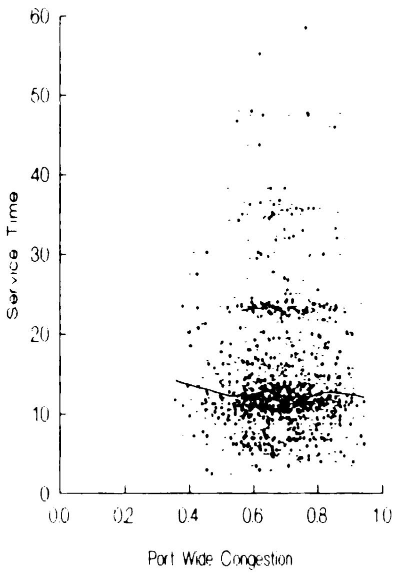


Berths 22-23: Service vs Congestion

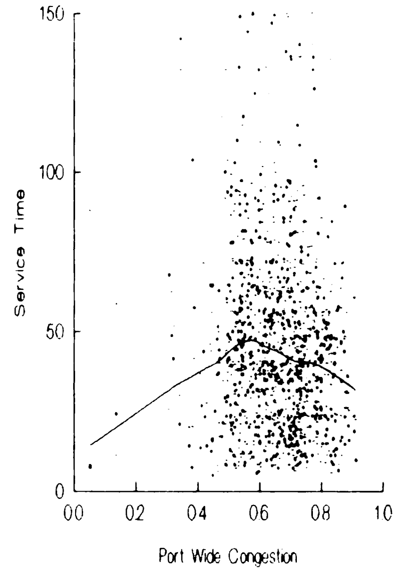


Berths 24-25: Service vs Congestion

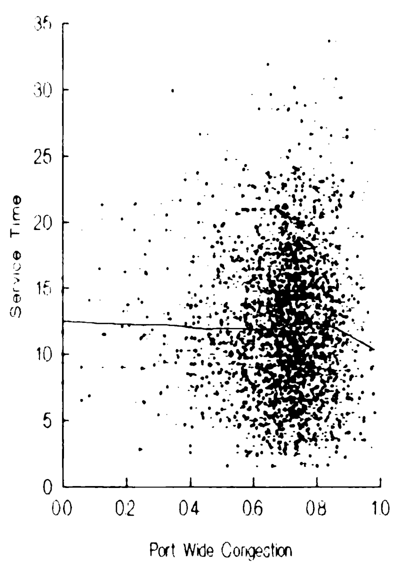
Figure 6
Lowess Plots: Service Time vs Port Wide Congestion for All Berths



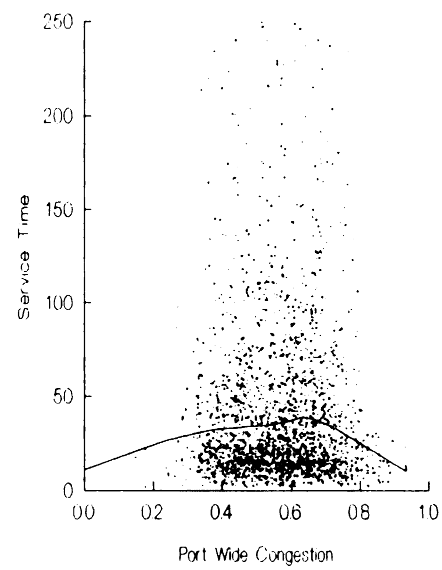
Berths 1-2: Service vs Port Wide Congestion



Berths 5-7A: Service vs Port Wide Congestion

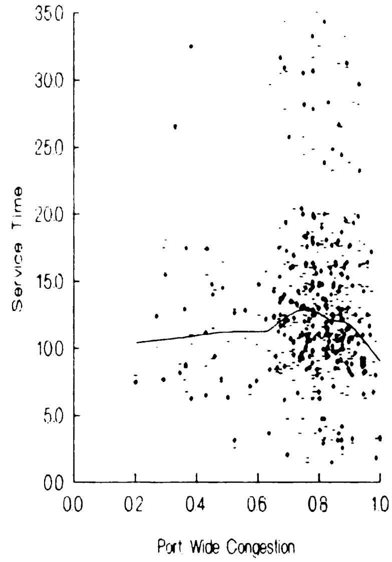


Berths 8-11: Service vs Port Wide Congestion

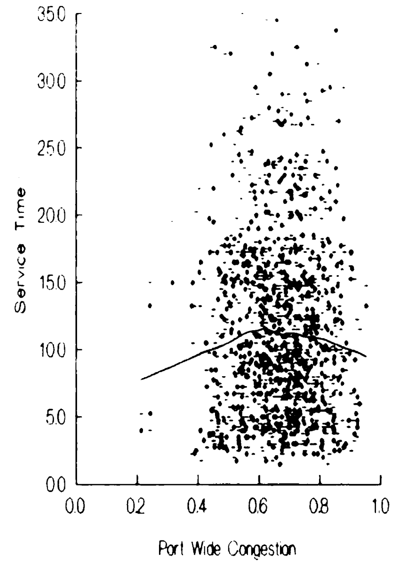


Berths 12-18: Service vs Port Wide Congestion

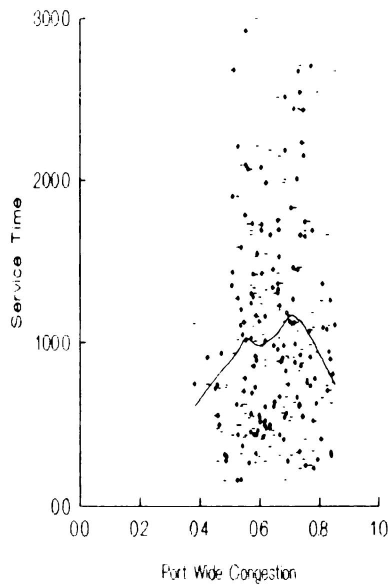
Figure 6 (Cont'd)



Berths 19-21: Service vs Port Wide Congestion



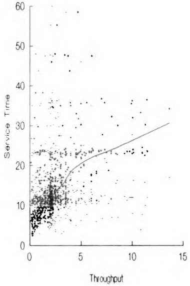
Berths 22-23: Service vs Port Wide Congestion



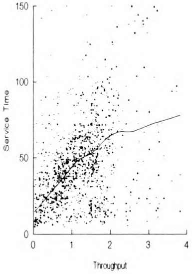
Berths 24-25: Service vs Port Wide Congestion

Figure 7

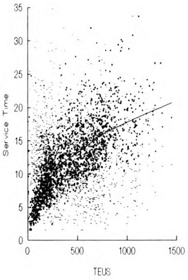
Lowess Plots: Service Time vs Throughput for All Berths



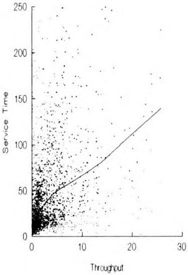
Berths 1-2: Service vs Throughput



Berths 5-7A: Service vs Throughput

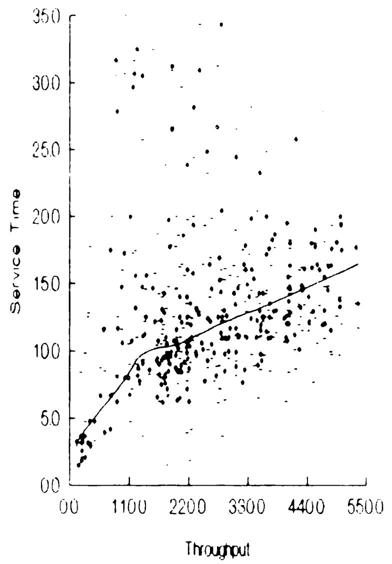


Berths 8-11: Service vs TEUS

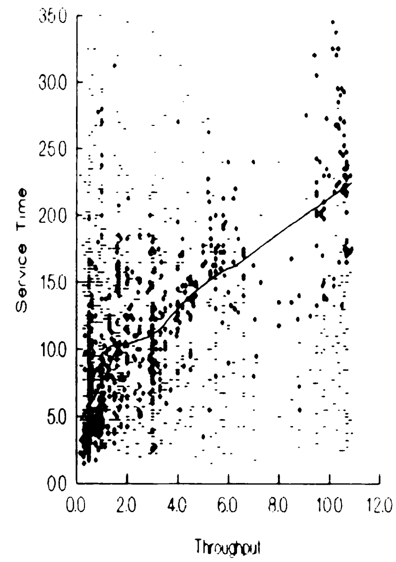


Berths 12-18: Service vs Throughput

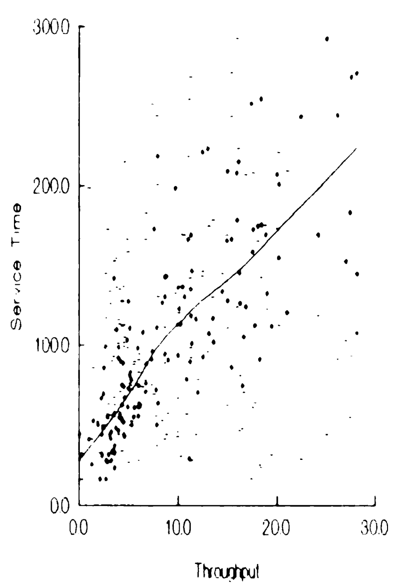
Figure 7 (Cont'd)



Berths 19-21: Service vs Throughput



Berths 22-23: Service vs Throughput



Berths 24-25: Service vs Throughput

Appendix 4

Appendix 4

The Bruesch-Pagan statistic is given by the following formula and has a Chi Squared distribution with r degrees of freedom where r is the number of variables that influence the error variance.

$$B = \frac{S_0}{2\hat{\delta}^4}$$

where S_0 = Regression sum of squares from the regression of residuals squared on the relevant variables.

$$\hat{\delta}^2 = \sum \frac{\text{Residuals squared}}{\text{Sample size}}$$

Table 21 below gives the results of the regression of residuals squared on variables that influence the error variance and the Bruesch-Pagan statistics for all subsystems.

Table 21

OLS of Residuals Squared and The Results for the Bruesch-Pagan Test

Berths 1-2

Variables	Coefficient	t stat	F stat	p value
Constant	0.28	9.29	14.40	0.000
Congestion	0.18	2.14		
Congestion Sqr	-0.19	-2.35		
Log Throughput	-0.03	-2.89		
Palmoil Cargo	-0.17	-5.92		
Petroleum Cargo	-0.25	-8.40		
Latex Cargo	-0.15	-3.49		
Chemical Cargo	-0.12	-2.70		

Berths 5-7A

Variables	Coefficient	t stat	F stat	p value
Constant	0.12	6.94	43.30	0.000
Log Throughput	0.11	4.89		
Log Throughput Sqr	0.07	5.21		
Standard Cargo	0.30	7.75		

Berths 8-11

Variables	Coefficient	t stat	F stat	p value
Constant	0.54	14.79	158.10	0.000
Log TEUS	-0.08	-12.57		

Berths 12-18

Variables	Coefficient	t stat	F stat	p value
Constant	0.46	10.38	7.60	0.000
Log Throughput	-0.06	-4.04		
Vehicles Cargo	-0.27	-3.96		
Ironstee Cargo	-0.01	-0.11		
General Cargo	-0.06	-1.23		
Chemical Cargo	-0.20	-1.98		
Palm Kernel Cargo	-0.27	-2.85		
Timber Cargo	-0.31	-4.42		
Sugar Cargo	-0.05	-0.47		

Berths 19-21

Variables	Coefficient	t stat	F stat	p value
Constant	0.76	4.88	14.70	0.000
Log TEUS	-0.11	-3.83		

Table 21 (cont'd)

Berths 22-23

Variables	Coefficient	t stat	F stat	p value
Constant	0.20	15.78	28.10	0.000
Log Throughput	-0.06	-5.30		

Berths 24-25

Variables	Coefficient	t stat	F stat	p value
Constant	0.312	6.984	20.0	0.000
Log Throughput	-0.09	-4.472		

Bruesch-Pagan Tests

Berths	Test Statistics	Degrees of freedom	Critical value at 5% level of significance
1-2	1532.73	7	0.99
5-7A	532.67	3	0.07
8-11	1909.65	1	0.0004
12-18	120.65	8	1.35
19-21	41.14	1	0.004
22-23	42.21	1	0.004
24-25	48.78	1	0.004

References

Allen, A., *Probability, Statistics and Queueing Theory*, 1990, Academic Press Inc., Boston, Massachusetes.

Arnott, R., De Palma, A., Lindsey, R. (1990). Economics of a bottleneck, *Journal of Urban Economics*. **27**, pp 111-130.

Ben-Akiva, M., De Palma, A., Kanaroglou, P. (1986). Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science*, **20**, pp 164-181.

Brown, G., Johnson, B., Public Utility Pricing and Output Under Risk, *American Economic Review*, 1969, **59**, No. 1, pp 119-128.

Cooper, L., Bhat, N.and Larry L., *Introductions to Operations Research Models*, 1977, W. B. Saunders Company, Philadelphia.

Cooper, R., *Introduction to Queueing Theory*, 1981, North Holland Publication, New York.

Dowd, T., Port Finances and Operations: Understanding the Bottom Line, in *Urban Port and Harbor Management*, 1988, Taylor and Francis, New York, pp 217-233.

Edelson, N., Hilderand, D., Congestion Toll for Poisson Queuing Processes, *Econometrica*, 1974, **43** No. 1, pp 81-91.

Expenditure Report, in *US Port Development*, Office of Port and Intermodal Development, 1991, US Department of Commerce, Washington DC., pp 40-41.

Gorney, L., *Queueing Theory: A Problem Solving Approach*, 1981, Petrocelli Book Inc., New York.

Gross, D. and Harris, C., *Fundamentals of Queuing Theory*, 1985, John Wiley and Sons, New York.

Jansson, J., Shneerson, D., *Port Economics*, 1982, The MIT Press, Cambridge, Massachusetts.

Kadir, E., Investment on Port Facilities- Too Much or Too Little, *Proceedings of the 1993 Transport Conference*, 1993, Chartered Institute of Transport, London, England.

Kleinrock, L., *Queueing Systems*, Vol 1, 1975, John Wiley and Sons, New York.

Knudsen, N., Individual and Social Optimization in a Multiserver Queue with a General Cost-Benefit Structure, *Econometrica*, 1972, **40**, pp 515-528.

KPA (i)- Commercial & Public Relations Dept., *The Story of Port Klang Malaysia*, 1988, pp. 6-7, Klang Port Authority, Malaysia, Klang, Malaysia.
KPA (i) - Commercial & Public Relations Dept., From Railway Port to Privatized Port, *The Gateway to Malaysia*, December 1992, pp. 9-10, Klang Port Authority, Malaysia, Klang, Malaysia.

KPA (ii) - Commercial & Public Relations Dept., The General Manager On KPA After Privatization, *The Gateway to Malaysia*, December 1992, pp. 9-10, Klang Port Authority, Malaysia, Klang, Malaysia.

KPA- Commercial & Public Relations Dept., Notes to the Accounts, *Port Klang Authority Annual Report*, December **1991**, pp. 45-64, Klang Port Authority, Malaysia, Klang, Malaysia.

KPA (i) - Commercial & Public Relations Dept., Notes to the Accounts, *Port Klang Authority Annual Report*, December **1994**, pp. 2-3, Klang Port Authority, Malaysia, Klang, Malaysia.

KPA (ii) - Commercial & Public Relations Dept., Notes to the Accounts, *Port Klang Authority Annual Report*, December **1994**, pp. 39-44, Klang Port Authority, Malaysia, Klang, Malaysia.

Lippman, A., Stidham, S., Individual Versus Social Optimization in Exponential Congestion Syatem, *Operations Research*, 1977, **25**, pp 233-247.

Mohring, H. (1994). *The Economics of Transport, Vol. 1.* Aldershot, England: E. Elgar Publication.

Naor,P., The Regulation of Queue by Levying Tolls, *Econometrica*, 1969, **37** No. 1, pp 15-24.

Newell, G., *Aplications of Queueing Theory*, 2nd Edition, 1982, Chapman and Hall, London, England.

Nicolaou, S., Berth Planning by Evaluation of Congestion and Cost, *Journal of Waterways and Harbors Division*, **93**, November 1967, pp 107-132.

Plumlee, C., Optimum Port Size, *Journal of Waterways and Harbors Division*, **92**, August 1966, pp 1-24.

Ross, S., *Probability Models,* 1993, , Academic Press Inc., Boston, Massachusetes.

Rue, R., Some Properties of Optimal Control Policies for Entry to an M/M/1 Queue, *Naval Research Logistics Quarterly,* **28**, December 1981, pp 525-532.

Robinson, R., Size of Vessels and Turnaround Time, *Journal of Waterways and Harbors Division*, **104**, May 1978, pp 161-178.

Ross, S., *Probability Models,* 1993, , Academic Press Inc., Boston, Massachusetes.

Schulze, G., Investigation and Review of Periodical Fluctuations in Port Turnovers, *Economic Computation and Economic Cybernetics Studies and Research*, 1978, **12**, No 3, pp 59-67.

Sherman, R., Visscher, M., Second Best Pricing with Stochastic Demand, *American Economic Review*, 1978, **68**, No 1, pp 41-53.

Taylor, H. and Karlin, S., *An Introduction To Stochastic Modeling*, 1994, Academic Press Inc., Boston, Massachusetes.

Tirole, J, The Firm as a Long-Run Relationship, *The Theory of Industrial Organization*, 1988, pp. 21-29, MIT Press, Cambridge, Massachusetts.

Wanhill, S., Further Analysis of Optimum Size Seaport, *Journal of Waterways and Harbors Division*, **100**, November 1974, pp 377-383.

Walters, A., Bennathan, E., *Port Pricing and Investment Policy for Developing Countries*, 1979, Oxford University Press, Oxford, pp 31-42.

Weille, J., Ray, A., The Optimum Port Capacity, *Journal of Transport Economics and Policy*, September 1974, pp 244-259.

World Trade Pattern in *World Economic and Social Survey*, 1994, United Nations, New York, Table 111.1, p 63.

Yechiali, U., Customer's Optimal Joining Rules for the GI/M/s Queue, *Management Science,* **18**, 1972, pp 434-443.

Yechiali, U., On Optimal Balking Rules and Toll Charges in the GI/M/1 Queuing Process, *Operations Research,* **19**, 1971, pp 349-370.