2020

A Theory of Fire Service Provision: With an empirical analysis of response time, suppression time, and service output

Hyong-gu Hwang
The College of Wooster, hhwang20@wooster.edu

A Theory of Fire Service Provision:
with an empirical analysis of response time, suppression time, and service output

Independent Study Thesis

Presented in Partial Fulfillment of the
Requirements for the Degree Bachelor of Arts in
the Departments of Economics and Mathematics
at The College of Wooster

by
Hyong-gu Hwang

The College of Wooster
2020

Advised by:

Dr. Jim Burnell (Economics)

Dr. Marian Frazier (Mathematics)

Abstract

We introduce a two-stage theoretical framework of fire services that justifies the status of response time as a factor input. In the first stage, the provincial government distributes a budget to its cities, resulting in varied numbers of firefighters and fire engines in each city. In the second stage, each city fire department places its fire stations at spatially optimal locations that minimize expected response times. When a fire occurs, the outputs from these two stages are actualized into dispatch level, response time, and suppression time. These intermediate outputs are then transformed into inputs for producing service output, which is measured in terms of fire spread. Using a data set of 49,000 fire dispatches that occurred in Gyeonggi Province, South Korea in 2014-2018, we estimate a set of models for the above outputs. We find evidence for increasing returns to population scale, while empirically showing that response time and suppression time are indeed inputs for the production of fire services.

For My Father

Acknowledgements

I would like to thank my advisors for their excellent guidance and support. Dr. Burnell has engrained in me the importance of context and linkage in academic research; Dr. Frazier set an example of a patient teacher and an intelligent communicator. Dr. Jia was my first mentor for this research and I am lucky to have her as my friend.

Special thanks go to Dr. Jaldell at Karlstad University in Sweden, who is a pioneer and an expert in the productivity studies of fire services. I enjoyed the privilege of asking him questions and receiving his feedback. Any errors or mistakes in this paper are of my own and not of my advisors nor Dr. Jaldell.

The Henry J. Copeland Fund enabled me to travel to South Korea, acquire data, and interview fire experts in 2019. The McDonald Fund and APEX Travel Grant allowed me to present my poster at the Mathematics Association of America conference; and the Jenny Research Fund provided the opportunity to present my paper at the Eastern Economics Association conference this year.

My college education is entirely due to the support of my parents and my grandma. Their love and sacrifice have always been the reason I put my best effort in everything I do. Lastly, thanks Jeeho, for being my one and only sister.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Unlike manufactured goods, output in fire services can be difficult to quantify due to its intangible nature. In fact, it may be difficult to conceptualize, let alone quantify, output to begin with. What exactly is being *produced* in the case of fire services? In what form does output exist? How do we measure it? In answering these questions, one may envision a fire station equipped with firefighters and fire trucks. A fire station as a **physical building** is certainly one form of output, but it fails to provide a complete picture of the public good. Some fire stations are busier than others, implying that the *quantity* of service provided may differ even if the size of the station is the same.

Following this train of thought, one may proceed to consider indicators of a station's **busyness**, such as the total number of fire dispatches per month or total fire suppression time. These indicators contain information about the amount of work performed by different fire services. However, defining output becomes even more challenging if one considers that fire service has a *quality* dimension. **Promptness** of response is an integral component of a fire service because fire, in many cases, is an emergency situation. Fire can spread over time, making the delivery of service sensitive to time. Meanwhile, from an economic standpoint, one may associate output with **value saved**. After all, saving the lives and property of citizens is the *raison d'être* of fire services. Considering this end goal of the service, it seems reasonable to define output in terms of the outcome of a fire.

As such, defining output in the case of fire services can be a challenging task because there are multiple aspects to be considered. Existing economic studies of fire service production have defined output in largely two distinct ways[1]. The first camp of researchers consider output as the **sum of the inputs** that go into service provision (Hirsch, 1959; Ahlbrandt, 1973; Hitzhusen, 1973; Kristensen, 1983). For them, firefighters and fire engines simply *constitute* output. As a result, output is measured in terms of the costs of labor and capital, which are revealed in the size of government expenditure. The studies aim to identify the factors that "affect" expenditure, such as population, area, density of houses, and real estate values, although the selection of control variables may vary, ranging from urbanization index, insurance costs, percentage of commercial properties, environmental variables, and so on. However, their focus is commonly set on finding the factors that determine the size of a fire service.

The second camp defines output in terms of the **consequences** of a fire, such as casualties (Jaldell, Lebnak, and Amornpetchsathaporn, 2014), dollar damage to property (Coulter, 1979; Duncombe, 1991; Duncombe and Yinger,

[1] Of course, there are also a number of studies that have defined output in ways other than the two that we have identified. In fact, Jaldell (2002) provides a thorough review of literature and categorizes the definitions of output into five categories. Readers interested in the details of this topic may find the first chapter of the author's work useful.

1993), square feet measure of physical damage (Ignall, Rider, and Urbach, 1979), and degree of fire spread (Jaldell, 2005; Jaldell, 2019). The advantage of this approach is that the quality or effectiveness of the service provided is taken into account. For citizens, the consequence of fires may be of greater importance than the sheer number of fire stations within the city. For this reason, Bradford, Malt, and Oates (1969) refer to the consequence of a fire as the outcome of interest to voters. Given that the ultimate goal of fire services is to protect the lives and property of citizens, indicators of consequence provide a more vivid picture of what is being produced by the fire services.

This paper aims to integrate the two camps of thoughts into a single theoretical framework. This integration necessitates a distinction between the *ex ante* and the *ex post*. In a sense, the focus of the first camp is more heavily set on determining the size of a fire service *before* a fire occurs. On the contrary, the consequence of a fire, by definition, exists only *after* a fire occurs. Thus, the second camp automatically situates itself in an *ex post* framework. The cost functions developed by the first camp will fit into what we will define as the first stage of *ex ante* production: the distribution of provincial budget across cities according to varied fire risks faced by each city. However, this process in itself is not our primary focus and we will not build or estimate any models for government expenditure.

Instead, we attempt to model a production function for fire services using information about the *consequence* of fires. That is, we follow the output definition of the second camp. However, there will be two notable differences. First, the first-stage distribution activity will be an integral part to constructing our output. In fact, the first-camp output will be defined as an *intermediate* output that eventually becomes an input for the second-camp output. This will be where the integration of the two camps occurs. Second, we will develop a theory that justifies the status of response time as a factor of production. This will entail defining the second stage of *ex ante* production: the spatial allocation of fire stations by the city fire department.

Moving on, the two-stage procedure of the *ex ante* production will be extended to the *ex post* production of fire services. We believe that this is a contribution to the current state of literature. Existing studies of fire services do not consider the linkage between the *ex ante* and *ex post* production activities. Recognizing this linkage allows one to have a better picture of how fire services are organized and provided. It also offers a theoretical validation of econometric models that treat response time as an input for fire service production (Wallace, 1977; Ignall, Rider, and Urbach, 1979; Morley, 1986; Jaldell, 2002; Jaldell, 2005). This validation may be useful because time-oriented factors are not conventionally viewed as factors of production, compared to standard inputs such as labor or capital.

So far, we have discussed how past studies have defined output in the case of fire services. Then, we presented the structure of our theoretical framework and explained where this paper stands within the current state of literature. Now, we conclude with an overview of the chapters to follow.

In Chapter 2, the details of our theory are presented. As we enter the *ex post* framework, we define the intermediate outputs as well as the final version of the output, which we call the service output. As a result, we will be able to explain the relationships between different layers of outputs and their relevant inputs in a systematic manner.

In Chapter 3, we review three empirical papers that are important to this study. Particular emphases will be made on: the conceptual framework, output definition and proxy variables, data set and nature of the sample, key components of the econometric models, and important findings. All these aspects of the papers will be discussed in connection to the current paper.

In Chapter 4, we introduce the three statistical methods that were considered for this paper: linear regression using Ordinary Least Squares estimation, instrumental variable method using Two-Stage Least Squares estimation, and ordered probit model using Maximum Likelihood estimation. We will explain the underlying logic of each method with an example.

In Chapter 5, we specify our models for the three *ex post* intermediate outputs and the service output. We aim to establish a causal relationship between the explanatory variables and the dependent variables of each model. Then, we introduce the composition of our data set. Lastly, we explain how the dependent variable for the service output model was constructed.

In Chapter 6, we present the estimation results for the following models: response time, suppression time, and service output. We aim to provide meaningful interpretations of the results, while also pointing out the limitation of each model.

Finally, Chapter 7 concludes.

Chapter 2

Economic Theory

In this chapter, the production of fire services is described as a two-stage process in which the economic agents of each stage engage in a sequential optimization. This two-stage theoretical framework is inspired by Jaldell (2002) who, to our knowledge, first defined response time as an intermediate output and also as an input for producing fire services, rather than a simply exogenous factor. We consolidate this argument by identifying the source of this intermediate output. That is, in order to justify the status of response time as an output, a relevant production activity needs to be defined. We formalize this argument by extending the theoretical presentations of Duncombe and Yinger (1993).

Section 2.1 provides a brief background of how fire services are organized and operate, for readers who are not familiar with the entity. This discussion will be made in the context of the jurisdictional structure of South Korea.

Section 2.2 introduces the two-stage optimization problem by defining the economic agents and their decisions carried out in each stage. The type of strategic interdependence is assumed to be sequential and not simultaneous;

that is, the first-stage decision affects the second-stage decision but not vice versa. We will state the advantage of this framework and refute two possible objections.

Section 2.3 defines the intermediate outputs of the two-stage decisions in an *ex ante* framework, meaning that the decisions are made prior to the incidence of a fire. In Section 2.4, these concepts are extended to an *ex post* framework where we now consider individual fire dispatches *after* the incidence of fires. As a result, we will be able to define response time and suppression time as inputs for producing fire services.

Lastly, Section 2.5 will present a model of average response time (Kolesar, 1975). Although the model will *not* be empirically tested in this paper, it will help understand the relationship between response time, number of fire stations across a city, and area. In particular, we will introduce the concept of service reach, which will appear later in our model specification.

2.1 Background

The organizational structure of South Korean fire services is as follows. A **province** consists of cities and sub-regions; each **city** has its own fire department; and fire stations are located in a subset of **sub-regions** across the city. Most sub-regions automatically fall under the jurisdiction of a particular city. There exist extremely large sub-regions that have its own fire department; these will be referred to as a city for the sake of convenience. Such sub-regions are very rare and their status are typically raised to a city over the years. Fire stations are placed in a subset of the sub-regions but not necessarily in all of them. The province is the entity that independently organizes and provides funding for its own fire services; sub-regions do not have the ability to collect taxes or organize their own fires services (unlike municipalities in the case of United States).

A **city fire department** acts as the headquarters for its fire stations and substations within a city. The city fire department building is also where all the administrators and higher-ranked officers are located. Therefore, it is usually located in the center of the city to provide citizens with better access to public services. The city fire department also tends to be equipped with specialized vehicles, such as rescue trucks or ladder trucks, as it is located in the most urbanized area within the city where these vehicles are most likely needed. On the other hand, **fire stations** are located in different parts of the city to meet the demand for fire service across the city. All fire stations are equipped with a pump truck and a tank truck. Depending on the level of fire risks within the neighborhood, additional vehicles are often allocated to fire stations as well. **Substations** are placed in areas with very low population density and are usually equipped with one pump truck. All fire departments have multiple paramedic and one rescue squad across the city, each of which has one or two ambulances and one or two rescue trucks. These squads operate as a separate entity from the fire stations but are typically located in the same building with firefighters for management purposes.

2.2 A Two-Stage Sequential Process

The production of fire service is a sequential process that occurs in two stages. In the first stage, the provincial government makes the **distribution** decision of how much budget should be spent for each of its cities. Cities have different population density, geographical characteristics, and number of fire-hazardous facilities, which are all determinants of the level of fire risk. The provincial government evaluates these risk factors and distributes fire resources (fire stations, fire trucks, and firefighters) to the cities. In the first stage, the goal is to acquire a desired level of social outcome using minimum amount of budget. In other words, the objective is to minimize the cost of producing a certain level of output.

In the second stage, each city fire department makes the **allocation** decision of where to place its fire stations. Fire stations are established at the spatially optimal locations so that response time is minimized. Thus, the second-stage objective is to maximize output, in the form of minimized response time, given the budget constraint predetermined in the first stage. In both stages, legal regulations are considered as they specify the minimum number of firefighters and fire trucks required for a given population, area, and number of fire-hazardous facilities. Although the terms 'distribution' and 'allocation' are often used interchangeably in other contexts, each will specifically and exclusively refer to the relevant stage of production activity throughout this paper.

2.2.1 Advantages of the Framework

Existing economic studies of fire service limit the definition of production only to the distribution decisions carried out by provincial governments. However, the allocation decision clearly impacts the quality of service provided and thus should be considered a part of the production process. The inclusion of the allocation decision into the production process has practical value. To demonstrate this point, let us assume that a city fire department identifies inefficiency in the operation of one of its fire station. This inefficiency may be revealed in prolonged response times which would lead to increased damages from fire. Without consideration for the allocation decision, the only argument that the city fire department can make is that it needs to receive more budget from the provincial government. However, with consideration for this issue, the city fire department can now identify the source of inefficiency more accurately. If the inefficiency is due to sub-optimal placement of fire stations, the city fire department may choose to reallocate its stations in different locations, as long as the benefit of doing so exceeds its costs. If the allocation proves to be optimal, then this justifies the need for more fire budget endowed from the provincial government. In both cases, consideration for the two stages will enable the city fire department to achieve greater social welfare within a neighborhood or the city.

2.2.2 Possible Objections

One may object to the claim that spatial allocation of fire decisions counts as a production activity. One possible argument is that response time is only a *quality* indicator and has nothing to do with the *quantity* of service produced. This claim is certainly consistent with the neoclassical expression of production where quantity produced is a function of capital and labor inputs. However, we will demonstrate that the quality dimension is integral to output measurement in the case of fire services considering the nature of what is being produced as well as the objective function of the producer.

Another objection may be that response time is only an exogenous variable. That is, one may acknowledge that response time affects the quantity of output but argue that firefighters do not have any control over response time. According to this argument, response time cannot be defined as a factor input because the producer does not have control over it. However, we will show that response time exists as a result of a city fire department's decisions and is thus fundamentally within the control of the producer of the service.

In the following section, we shall develop a theory of fire service production by carefully defining input, output, and outcome. In particular, we will newly introduce spatial allocation and response time to the theoretical framework of Duncombe and Yinger (1993). This will provide a more comprehensive understanding of the organization and provision of fire services.

2.3 A Theory of Fire Service Production

Following the definitions presented in the text, *Microeconomics: Theory and Applications with Calculus by Jeffrey M. Perloff (2009)*, we define **output** as the goods and services that a firm sells or provides; and **inputs** as factors of production such as labor, capital, and material that are used to produce an output. In this study, we narrow the scope of fire services to firefighting activities only, while excluding fire prevention and administrative duties from consideration. Then, output can be broadly defined as the sum of all firefighting activities produced by the provincial government and provided through the agency of the city fire department. As suggested in the previous chapter, the *ex ante* and *ex post* production should be considered separately in order for one to better understand exactly *what* is being produced. Before a fire actually breaks out, the production of fire services occurs as a sequential process of two stages, each of which results in different types of output. After firefighters respond to a fire, *ex ante* output is transformed into *ex post* output under the influence of environmental variables and fire conditions.

2.3.1 First-Stage Output: Stand-by Level

In the first stage of production, the provincial government produces the availability of city fire departments of different sizes across its cities. That is, the provincial government provides citizens with the availability of fire departments and the subsidiary fire stations that are equipped with the adequate amount of fire resources given the environment. Determination of the adequate requires evaluation of varied risks of fire in each neighborhood, which are reflected in factors such as population density or number of hazardous facilities. For example, neighborhoods with higher population density are given a larger number of firefighters compared to less populated ones. Also, ladder trucks are distributed to neighborhoods with tall buildings, but not to areas without vertical structures. In both cases, different amounts of fire resources are distributed to the neighborhoods so that firefighters can readily respond to fires with the appropriate equipment.

We notice that the first-stage output has two aspects. The first aspect is the *size* of fire stations, which is determined by the size of the budget distributed to each city from the provincial government. The second aspect is *availability*.

The provincial government does more than merely purchasing fire trucks and hiring firefighters; it ensures that fire service is provided *whenever* citizens need it. In other words, the first-stage output is produced no matter what happens. These two aspects can be encapsulated into the term **stand-by level** (Jaldell, 2002). Following Duncombe and Yinger (1993), stand-by level can be expressed as follows:

$$G = g(L, K, Z) \tag{2.1}$$

where L is labor, K is physical capital and equipment, and Z is all 'other inputs' such as GPS technology specialized for fire services. The function g is similar to the production function of a private firm because it defines the way in which inputs are combined to produce an output. In the meantime, G can also be defined as the level of governmental activity because it is directly proportional to the size of the budget endowed from the provincial government. However, G only reflects a particular size of a fire service and its constant availability to citizens, without any consideration for the actual location of the fire service. In other words, fire departments and fire stations do not physically exist at this stage.

2.3.2 Second-Stage Output: Firefighting Power

In the second stage of production, the city fire department places its fire stations at optimal locations so as to minimize expected response time, or the expected time it takes a fire service to arrive at the fire scene. When a fire incident occurs, firefighters and fire trucks are sent to the scene to suppress the fire as promptly as possible. In economic terms, these firefighters and fire trucks are the inputs, while their capacity to suppress the fire is the output at this stage of production. We will call this capacity the **firefighting power** of a fire service. However, labor and capital are not the only factors that determine firefighting power. Response time is also a crucial factor because a fire typically aggravates over time and prompt arrival can change the course of the fire. Given the same severity of fire and same number of firefighters and fire trucks involved in a dispatch, a shortened response time will enhance the firefighting power of a fire service. For this reason, expected response time should be considered an input that determines the level of firefighting power. Therefore, the second-stage output can be expressed as:

$$F = f(G, A) \tag{2.2}$$

where A is the vector of decisions relating to the spatial allocation of fire stations by the city fire department. For average-sized fire incidents, the firefighting power of a fire department would be a vector of the firefighting power of the subsidiary fire stations. However, in some cases, the collective firefighting power of a city fire department may be greater than the sum of its parts (i.e. fire stations). For example, a massive fire may occur in which multiple fire stations become involved, or fires may occur simultaneously within a neighborhood and require one fire station to cover for another. In these cases, proximity between fire stations will produce a synergy effect that is greater than the mere sum of each fire station's firefighting power. As such, the specific function of a fire department's firefighting power may differ across cities. Thus, f should be understood as a generalization of such functions.

2.4 *Ex Post* Output

When a fire incident occurs, firefighters and fire trucks closest to the location are dispatched to the fire scene. In the *ex post* framework, it is reasonable to make the following assumptions. First, firefighters always 'do their best' to suppress a fire. In other words, assuming identical fire severity, the level of suppression effort is the same regardless of the value of the property on fire. Second, firefighters always aim to arrive at the fire scene as promptly as they can[1]. Exogenous factors such as weather, road, or traffic conditions may yield some variance in response times. However, response time is predominantly determined by the location of the fire station, which is the product of the second stage of service production.

The sequential procedure of the *ex post* production largely resembles its *ex ante* counterpart. Depending on the severity of a fire, only a subset of a fire station may be dispatched to the scene. In other words, not all firefighters and fire trucks may respond to suppress the fire. Let L', K', and Z' each denote a subset of firefighters, fire trucks, and other fire resources that are involved in a particular fire incident. Then, these subsets of first-stage inputs would produce a certain level of first-stage output. We will call this *ex post* first-stage output the **dispatch level** and denote it as G'. While dispatch level is analogous to the stand-by level, G, it is different in that only a subset of fire resources are actually dispatched to the scene. Thus, dispatch level can be written as follows:

$$G' = g(L', K', Z'). \tag{2.3}$$

[1]We will release this assumption when we specify the model for response time.

In the *ex post* framework, the prior spatial allocation decision, A, is actualized in the form of **response time**, R. It is important to recognize that both dispatch level and response time are predetermined by the two stages of *ex ante* decisions. In other words, the first-stage distribution decision and second-stage allocation decision have a bearing on the *ex post* production of fire services.

Given a certain amount of dispatch level and response time, ***ex post* firefighting power** is produced in a particular fire incident. Denoted F_{post}, this firefighting power is also analogous to its *ex ante* counterpart and can thus be expressed as:

$$F_{post} = f(G', R). \tag{2.4}$$

Here, F_{post} indicates the *maximum* potential of suppression activity that can be performed by a fire service. In other words, F_{post}, by itself, does not provide information about the specific level of suppression effort that was actually made to put out the fire. Whether firefighters utilize their firefighting power up to their full potential depends on the severity and conditions of the fire. For example, the actualized level of effort to suppress a fire in a barren field will certainly be different from that of suppressing a house fire, even if the same number of firefighters and same types of fire trucks were involved. The varied levels of effort may be reflected in the time it takes to suppress the fire.

In response to a fire incident, firefighters provide their firefighting power to the citizens for a certain amount of time, which will be referred to as **suppression time** and denoted T. The amount of *work* that is being done during this time is the very *output* that we have been trying to define; this is the final form of output.

We are now able to define **service output**, Q, as the amount of firefighting power that is produced when a group of firefighters and fire trucks respond to an emergency call and suppress the fire for a certain amount of time. Notice that this definition involves all of the intermediate outputs that we have established within the *ex post* framework: dispatch level, response time, and suppression time. To sum up, service output can be expressed as follows:

$$Q = q(L', K', Z', R, T). \tag{2.5}$$

As a result, we have established response time and suppression time as *outputs* of intermediate production activities as well as *inputs* for producing service output.

2.5 A Theory of Response Time

In spatial economics, **market reach** refers to the minimum area that a firm needs to serve in order for its service to be worthwhile (Christaller and Baskin, 1966). Although fire services are not organized in response to a market-oriented demand for emergency services, per se, we may consider the concept of market reach in the context of fire services. Here, we define the **service reach**, SR, of a fire station as follows and use the term in the remainder of the paper:

$$SR = \frac{A}{n} \tag{2.6}$$

where A is the area of a city and n is the number of fire stations in the city.

Response time will increase as service reach increases, as a fire station will have to cover a larger area of land and travel longer distances. Meanwhile, response time will decrease as the number of fire stations within a city increases; each fire station will cover a smaller service reach, as revealed in Equation 2.6.

We conclude this chapter with a fundamental model of response time that has been used in many applications. This model can be understood as an extension to the notion of service reach with a consideration for the number of "busy" fire stations. Following Kolesar (1975), expected response time (travel time), ET, can be expressed in the following relationship:

$$ET = \alpha + \beta \cdot \left[\frac{A}{n - \lambda ES}\right]^{\gamma} \tag{2.7}$$

where A and n follow the same definitions from Equation 2.6, λ is the expected number of fire dispatches per hour for all fire stations in the city, and ES is the expected suppression time (in hours). The specific values of the parameters α, β, and γ are dependent upon the physical characteristics of the city. The term, λES, is an approximation for the expected number of fire stations that are *not* available. Therefore, the denominator, $n - \lambda ES$, indicates an approximation for the expected number of fire stations that are available during a given hour[2].

Equation 2.7 suggests that response time would increase as fires occur simultaneously within a given area. That is, as the number of unavailable fire

[2]These approximations are based on a separate model for distance. As such, the described model is *not* an expression for a set of physical laws; rather, it is an "approximation combined into another approximation", as the author clarifies. This model has been employed in numerous applications even in recent years (Taylor, 2015) and has been shown to be robust to different samples.

stations increases, the expected response time will increase as well. This implies an interdependence between fire stations; the availability of one fire station may affect the performance of another by prolonging its response time to a fire incident. For this reason, the density of fire stations should also be considered when modelling fire service provision. Population density may not suffice as an indicator because some cities may be in a relative shortage of fire stations even though its population is high. In this sense, service reach can be a more effective indicator of whether a "sufficient" number of fire stations are in place within a city.

Chapter 3

Literature Review

In Chapter 1, we provided a survey of economic literature on the provision of fire services. The focus was set on identifying different definitions of output and the relevant measurement variables that were chosen for empirical analysis. In this chapter, we review three pieces of empirical literature with an emphasis on their conceptual framework and selection of independent variables for model specification.

3.1 Duncombe and Yinger (1993)

The goal of the following article is to systematically define 'returns to scale' in public production. A system of four dimensions are proposed. The first three are the service quality, level of governmental activity, and population scales; the fourth is the economies of scope. The article presents a thorough, mathematical derivation of each dimension by considering a cost function for fire service production. Rather than mentioning the technical details, we will focus more on the logic behind the derivations.

Their strategy of obtaining the cost function is to take the 'inverse' of a production function. A two-stage framework is considered. The first stage concerns the level of governmental activities, which are revealed in government expenditures. The second stage involves a qualitative aspect of the service, which is directly experienced by citizens. The government activities are defined as an intermediate output; and the outcome of interest to voters as the final output.

Before we proceed, a clarification may be useful. The first-stage intermediate output is analogous to our own definition of stand-by level. In our theoretical framework, the size (and availability) of a fire service is determined by the first-stage budgetary decisions. However, the article "jumps" directly to the final output without considering varied response times or suppression activities. This difference stems from their purpose being the construction of a cost function. Their interest does not lie on the stand-by level itself but on the *cost* of maintaining a fire station. Meanwhile, our focus is set on the production function of the fire services, not on governmental entities.

As a result of this difference, the unit of observations are different. In our paper, we naturally consider individual fire dispatches to analyze the role and impact of input factors on our output. Duncombe and Yinger (1993), on the other hand, consider a sample of 188 municipal fire departments in New York State in 1984-1986. Their analysis does not involve time- or space-related factors.

The article defines the relevant inputs for each output. Government activities are defined as a function of labor, capital, and other materials; outcome of interest to voters is expressed as a function of the government activities, a jurisdiction's population, and (a vector of) environmental variables. Notice that the intermediate output now becomes an input for the final output. That is, government expenditure becomes a factor that decides the outcome that citizens experience. However, these are affected by how many people are served and how harsh the firefighting environment is.

Here, we notice that an analysis of fire production, whether it be on an aggregated government-level or on an individual fire-level, involves population and external factors. However, the population is only considered in the context of how many citizens are being served and how costly it is to do so; our motivation for considering population will be to examine the marginal returns to output (namely, response time) rather than marginal costs.

After a series of algebraic manipulations, the article derives the cost function. Then, it takes the partial derivative of the total cost with respect to different scale and scope dimensions. These partial derivatives are the mathematical definition of returns to scale. Thus, the goal of their empirical analysis is to estimate the values of these scale parameters. The rest of the analysis is not relevant to our paper.

The article provides two interesting points to consider. First, their measure of the final output is property loss, L, relative to the property value, V, within a community. This proxy, $\frac{L}{V}$ can be written as follows:

$$\frac{L}{V} = \frac{L}{F} \times \frac{F}{V},$$

where F represents fires. To directly quote their explanation, "a reduction in property losses can either come from a decrease in the number of fires (fire prevention) or in the loss per fire (fire suppression)." This simple equation summarizes the two key aspects of the fire service. If data for continuous dependent variable relating to monetary value are available, this can be a useful approach.

Second, the concept of returns to *population* scale is worth investigating. This can be an indicator of how efficiently a fire service is serving its citizen. Notice the use of the word 'efficient'. Generally speaking, the conflict of efficiency and equity in fire services boils down to the issue of population and land size covered by a fire department. From an efficiency standpoint, the size and location of a fire service should be determined in such a way that the largest population (for a given area of jurisdiction) can be served. However, this results in a possibly unevenly greater allocation of resources to densely populated, urban areas. On the contrary, rural areas will receive a smaller amount of resources and have fewer fire stations. This leads to prolonged response times in these areas, which often makes a big difference on outcome in emergency situations. For this reason, regulations in South Korea dictate both land- and population-oriented guidelines for the *minimum* number of fire stations to be established across jurisdictions. Of course, due to a limited amount of resources, the population-oriented guidelines practically determine the allocation of fire resources. In this sense, considering both land size (service reach) and population density will allow one to carry out a more "balanced" analysis on the production of fire services.

3.2 Jaldell (2005)

The primary purpose of the following article is to compare the productivity differences among Swedish fire services. To do so, the article first establishes a conceptual framework within which fire suppression activity is delivered to citizens. As shown in Figure 3.1, the article separates the production of fire services to intermediate steps. Note that the figure is directly copied from the article.

In the first step, resource allocation of the Swedish government results

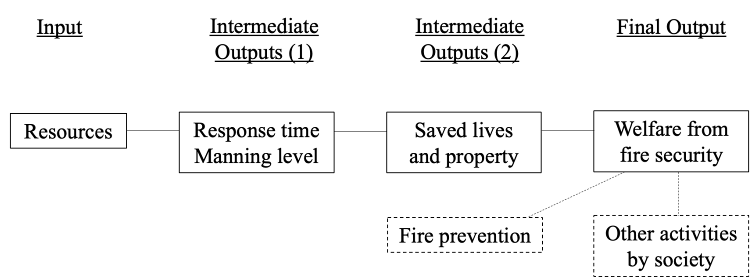

Figure 3.1: Flowchart of Fire Suppression Activity (Jaldell, 2002)

in varied average response times and manning level (number of firefighters initially reaching the fire). For the sake of consistency, we will use the term dispatch level instead of manning level. Response time and dispatch level are "assumed" to be outputs. The way in which the author introduces these elements into the framework is not based on any prior economic decision or activity; rather, the author explains that the "intermediate output can be measured both by response time, the faster the better, and by how many firemen that will turn out, the more the better"[1]. In other words, the two are defined as

[1]This statement was directly quoted from the dissertation of the same author (Jaldell, 2002), which includes a more detailed narrative of the author's conceptual framework.

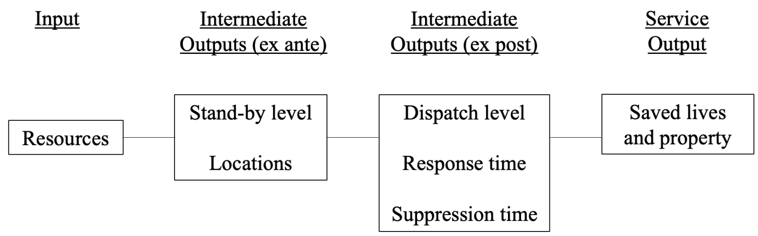

Figure 3.2: Revised Flowchart of Fire Suppression Activity (Hwang, 2020)

outputs because they are important *indicators* of fire suppression activity.

In the theoretical framework that we proposed in the previous chapter, we identified the very *source* of these intermediate outputs. As shown in Figure 3.2, dispatch level originates from stand-by level, which is determined in the first stage of the *ex ante* production: distribution of government budget. Meanwhile, we claim that response time originates from the specific location of a fire station, which is determined by the second-stage decision of the city fire department: spatial allocation of fire stations. In particular, we have established that the location decisions are *actualized* into response times when a fire actually occurs. Therefore, we believe that the revised conceptual framework in Figure 3.2 can function as a theoretical improvement to the framework in Figure 3.1; that is, the status of response time and manning level as intermediate outputs is validated through the identification of the relevant economic activities.

Jaldell (2005) points out that the performance of firefighters may differ even if the same size of dispatch level is sent out to the fire scene within the same response time. These productivity differences are reflected in the second intermediate output proposed by the article: saved lives and property. In producing this output, the first set of intermediate outputs are transformed into **inputs**.

This transformation justifies the inclusion of response time and dispatch level as **explanatory variables** within the econometric model for the second intermediate outputs.

As shown in Figure 3.1, the production scheme extends to the rather abstract notion of welfare from fire security, which is the final form of output. In producing the final output, the prevention aspect of the fire services needs to be considered. Since this is beyond the scope of our paper, we did not include this in Figure 3.2. Jaldell (2005) also does not perform any empirical analysis on the final output.

The measurement of the second intermediate outputs is the primary focus of the article. The article defines a criterion for an effective measurement: the proxy must capture the *difference* between the potential course of a fire and the result due to the firefighters' work. In this sense, the degree of **fire spread** is an effective proxy because it shows what could have happened and what was prevented by the fire service. However, the article acknowledges that the ideal proxy would have been the value of property and number of lives saved (in relation to the total the total value and number) and explains that relevant data were not available.

The author introduces a novel way of constructing the dependent variable. The author devises a measure of fire spread that compares the **fire condition upon arrival** and **where the fire was extinguished**. There exist multiple 'outcomes' based on this measurement, which the author categorizes into three ordinal levels. The details of this process will be presented in Chapter 4 of our paper, since we chose to use a similar proxy. As a result of this ordering process, the dependent variable consists of three levels that indicate better ($Y = 2$) to worse ($Y = 0$) outcomes.

Due to the dependent variable being discrete, an ordered probit model is used. The right-hand side of the regression equation takes the form of a typical linear model, consisting of five input variables and 37 individual city dummy variables. The input variables include: response time, number of own firefighters (initial crew), number of extra firefighters (additional crew), binary variable for full-time and part-time (firefighters), and another binary variable for life-saving activity. The city fixed effects are included for the purpose of productivity comparison among fire services. Out of 253 fire services, only the 37 with more than 20 dispatches are given a dummy variable.

The data set considered in the article consists of 3,039 residential fires that occurred across 288 Swedish municipalities in the years of 1996 and 1997. The sample only includes fires on detached houses, as they are understood to be the most homogeneous type of structure. The data set does not contain any information on environmental conditions or capital inputs, namely, fire engines and water equipment. While the author is correct in that firefighters and fire trucks are complements, we believe that fire trucks could be a better proxy, if data were available on both inputs. This would particularly be the case if there are multiple types of fire trucks. Including the number of fire trucks in the model would allow one to capture the different contributions that each type of truck makes. The number of firefighters would simply be a linear combination of the number of each truck.

The author restricts the maximum number of own firefighters to 20 and extra firefighters to 10. The reason is that each variable includes those firefighters who were put in to the scene after a shift change. That is, fires may continue at hours when the night-shift firefighters end their duty and the day-shift ones begin working. In Sweden, a shift change seems to occur, given that the author discusses an exchange of firemen at the scene. In South Korea, this does not happen; those who are sent to the fire dispatch complete the fire suppression even they go over hours[2]. Thus, we did not have to consider such a cap in our econometric model. However, our data set suffers the problem that all the initial crew and additional crew are added together without distinction. Therefore, we will also try the method of restricting the maximum number of the dispatch level, but our reason will be different from the article. Also, we will apply this method only for sensitivity analysis and not for our main model.

The binary variable for full-time firefighters is included to test the hypothesis that full-time professionals would produce better outcomes than part-time firefighters. The reasoning is that the full-time group receives better training and has more experience. In fact, their results show that there is no productivity difference between the two groups. The author thus infers that team spirit is a more decisive factor than the number or constituents of firefighters, at least in fires on detached houses. However, it is also possible that this result was due to the very small number of fires in which the initial severity was high. If their data set consisted of a larger number of harsh fires, then professional firefighters may have shown to be more productive. Our own data set shows a similar composition of severe and relatively trivial fires; the number of the former is disproportionately small.

Lastly, the life-saving dummy variable was included to test the hypothesis

[2]If a fire continues for days, which is not uncommon in large forest fires, then firefighters in South Korea do exchange shifts. However, in our sample of residential fires, no fires continued for over 6 hours.

that rescue activity would result in worse fire outcome. The reasoning is that rescuing human lives always comes first. Spending more time on the life-saving activity, fire suppression may be carried out at a slower pace or not carried out at all, thus allowing the fire to spread more.

In the described model, the role of the response time variable is rather questionable. While we fully agree with the motivation of including it as an input variable, response time should not have any impact on their outcome variable, given the way it was defined. Their dependent variable is designed to indicate the degree of fire spread between two time points: the time when the initial crew arrived and the time when the fire was extinguished. If a fire has already evolved to a greater magnitude, we are unable to know if this high initial severity of fire is due to prolonged response time or simply due to late discovery. Once a fire service arrives, the only factors of production that can make a difference in the outcome are labor and capital. Response time only matters in the sense that a fire can spread and aggravate the situation *while* the firefighters are traveling to the fire scene.

For this reason, we suggest that the definition of the dependent variable should rather involve the following time points: the time when an emergency call was made and the time when the fire was extinguished[3]. Under this definition, shortened response time directly impacts the fire outcome; it will mean that the initial crew arrives earlier, which will connect to a lower initial severity of a fire and possibly to better fire outcomes.

Finally, we discuss the fire service fixed effects. The author explains that the

[3]The obvious alternative would be to adhere to the original definition of the output and exclude the response time from the model.

37 dummy variables are analogous to the individual effects in panel data models. The author claims that the coefficient estimates on the fire service dummy variables can be interpreted productivity differences between the fire services. The article presents a table of these coefficient estimates and use this result to rank the productivity of each fire service. An analysis of this sort may serve as a useful decision tool for policy makers, as it enables a quantitative comparison between entities, whether it be individuals or fire services. However, one must also thoroughly examine any other factors that can enter the fixed effects.

By definition, fixed effects capture *all* the unobserved cross-sectional characteristics. The characteristics may be unobservable by nature; or they may be observable but relevant data may be unavailable. Although the cross sections in the above model are defined as fire services of each municipality, the fixed effects are essentially encapsulating all the characteristics of the *municipalities* and not only fire services. These characteristics may be strong enough to distort the impact of productivity differences within fire services. For example, what if certain municipalities mandated the installation of sprinklers in every single detached house? What if other municipalities had a larger proportion of old houses that were more susceptible to fires? Unless these factors are controlled for in the model, the possibility of these attributes entering the fixed effects remains. Even if all the relevant control variables are included, there still exist attributes that cannot be observed at all.

The model presented in the article does not include any control variables relating to external conditions. Given that the data were recorded in the late 1990's, a lack of detailed information is comprehensible. However, this very limitation may pose a question to the validity of fire service fixed effects as an indicator of a fire service's productivity. The following article to be presented mitigates this concern by including comprehensive and detailed information about each fire.

3.3 Jaldell (2019)

The following article is an extension of the author's previous work (Jaldell, 2005). The two purposes of the study are given as: to show how efficiency can be measured using a discrete outcome variable; and to measure efficiency for Swedish fire and rescue services and examine the source of these differences. The statistical method selected for this study is the logistic random parameter model. We will not discuss the technical details of the method, as our own model estimation will not involve random effects. Two versions of the dependent variable are considered: one being the three-level ordinal variable and the other being a binary outcome variable. The data set considered in the article consists of 29,813 residential fires that occurred in Sweden from 2009 to 2013, with 290 municipalities and about 150 fire and rescue services.

There are a few notable differences from the previous article. First, a binary version of the dependent variable is introduced. This may involve some aggregation of the data because the qualitative measure of outcome is now simplified, compared to its three-level counterpart. Instead, the interpretations of the parameter estimates may become relatively simple. Second, the sample considered has expanded from detached houses to residential fires. While the homogeneity of the observations has weakened, sample size has greatly increased. The author may have made this decision due to the availability of increased, detailed information about individual fires. The reason for converting from probit to logit models could not be found in the article.

The model estimation is performed in largely three steps. First, a similar model from the previous article (Jaldell, 2005) is estimated. However, only the response time variable is included as a key explanatory variable; instead, a long list of control variables are included in this model. The controls are categorized into three groups: fire suppression devices, reasons of fire, and starting room. Fire reasons include explosion, fire works, candles, and many other; starting room encompasses all spaces within a residential house, from bathroom and garage to stairwell and chimney. All of these controls are binary variables. Also, fire and rescue service (which used to be called fire service) random effects are estimated.

In the second step, the service-specific random intercept is used as a relative efficiency measure. In the third step, these random effects become the dependent variables, while the explanatory variables are: dispatch level (firefighters only), the previous full-time dummy variable, size of the service, and population of the municipality. The author states that this novel method of analysis is lacking in econometric foundation. For this reason, we do not employ this particular method in our model estimations. The more relevant aspect of this article to our paper would be the motivation for including the stand-by level and population variables.

The size of the fire and rescue service is included to observe any organizational differences that would connect to variations in productive efficiency. The reasoning behind this is that a larger service may offer better training opportunities and facilities. In South Korea, firefighters are employed by the provincial government and go through homogeneous training organized by the provincial fire academy. Therefore, we would not expect improved training opportunities, per se, in larger fire departments. However, fire departments in larger cities do seem to have more resources and may provide better facility. For example, the working conditions for firefighters may vary due to deteriorated capital equipment or poor training facilities.

In the meantime, larger population is explained to draw in more competent personnel both in fire crew and in management. South Korean firefighters are employed by the province not by municipalities, so we would not necessarily expect a firefighter in an urban area to be more competent than their rural counterpart. A more relevant explanation to our study may be in regards with population density. In densely populated areas, fire services will be able to cover a larger number of population within a shorter response time. To construct a more accurate indicator, we may also consider population density per fire station of a city. This will allow us to observe the impact that the density of fire stations have on response time or service output.

The article concludes with a few findings. First, a more "flexible" organization of fire services, such as one with first response persons, is said be more effective in fire suppression. First response persons, in Swedish fire services, are single firefighters whose role is to arrive at the fire scene two minutes before other services. Although such a concept does not exist in South Korean fire services, an ambulance may perform a similar role due to better mobility. The second finding is that populated municipalities are more efficient. However, this result is derived based on a variable that only indicates population. We claim that population density is a more relevant measurement, unless all municipalities are equal in land size. For our model specification, we will include the population density variable as well as population density per fire station. Since response time is dependent upon the spatial environments, a consideration for the land size of each city will be frequently made in our models.

Chapter 4

Statistical Methods

Regression is a powerful tool of analysis that allows one to quantify the impact of an independent variable on a dependent variable. The verb *regress* implies that the method returns data points back to a somewhat simplified form. Regression analysis allows one to identify patterns that exists within a data set and the relationship between variables. In particular, *linear* regression expresses the dependent variable as a linear function of one or more independent variables. The method is utilized for many analytical purposes such as creating a model that describes the reality; testing hypotheses about a theory or policy; or forecasting future outcomes by constructing predictive models.

We will begin this chapter by presenting the statistical theory behind linear regression analysis. Then, in section 3.2, we will introduce the concept of an instrumental variable, which is a method that has been increasingly used in econometric analyses. Lastly, in Section 3.3, we will describe the ordered probit model and its usefulness when the dependent variable is not continuous in nature.

4.1 Linear Regression

In order to demonstrate the underlying logic of linear regression, we begin with the simplest case of single-variable regression. We will define the key components of a regression equation, examine the derivation of coefficient estimates, and use an example to demonstrate how the results are interpreted.

4.1.1 Single-Variable Regression

A single-variable regression involves one independent (explanatory) variable affecting the dependent (outcome) variable. We may establish a 'theoretical' equation that perfectly describes our reality:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{4.1}$$

where ε_i is the **error term** which accounts for the stochastic variation in the dependent variable, Y. It is important to notice that the error term absorbs the randomness that cannot be encapsulated in our model of the real world. When real life data are collected, not all data points lie on the regression line. If all our observations lie on the line of the theoretical equation, then this implies that there was nothing for us to model in the first place. In such case, the relationship between X and Y are said to be purely deterministic and thus no estimation is required.

Equation (1) can also be understood as the relationship between X and Y within the entire population. In statistical inference, it is implicitly assumed that a complete set of data for the population is unavailable; otherwise, there would be nothing for us to infer because we already have full knowledge of the population. Thus, we use available data and produce an estimated equation for a given sample as follows:

$$\widehat{Y_i} = \widehat{\beta_0} + \widehat{\beta_1} X_i \tag{4.2}$$

where the subscript i represents the observation number ranging from 1 to N, and N is the number of observations in the sample. Deriving this equation is equivalent to finding the line that best describes the data points in our sample. In other words, we want to estimate the coefficients so that the estimated regression equation "fits" our data well. In obtaining the values of the coefficients, we may choose to use an estimator called Ordinary Least Squares.

4.1.2 Ordinary Least Squares

The ultimate goal of linear regression analysis is to create a regression equation by estimating the coefficients. In statistical terms, these coefficients are the parameters of interest, or quantities we wish to estimate. An **estimator** of a parameter is some function of the data that enables us to derive a value that is close to the true value of the parameter. This 'true' value is only theoretical and is usually not revealed in empirical studies, so the goal of the researcher is to choose the most effective estimator available. The one that is most widely used in linear regression is the Ordinary Least Squares (OLS) estimator.

OLS is a function that minimizes the sum of the squared residuals. A **residual** refers to the difference between the actual and estimated values of Y, in other words $e_i = Y_i - \widehat{Y_i}$. In single-variable regression, the residual can also be expressed as $e_i = Y_i - \widehat{\beta_0} - \widehat{\beta_1} X_i$ from Equation 4.2. As residuals are signed differences, adding them up is not desirable because distances of opposite signs may cancel out, which may not allow the information about the distances to be preserved. For this reason, we square the residuals. Therefore, the OLS estimator derives the line of best fit in a manner that minimizes the squared distance between data points and estimated values. In other words,

$$\text{OLS minimizes} \sum_{i=1}^{N} e_i^2. \tag{4.3}$$

Equation 4.3 is equivalent saying that OLS minimizes $\sum_{i=1}^{N}(Y_i - \widehat{\beta_0} - \widehat{\beta_1} X_i)^2$. Specifically, the OLS procedure calculates the estimates of our regression coefficients by the following formulae:

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}$$

and

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N}[(X_i - \bar{X})(Y_i - \bar{Y})]}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

where $\bar{X}$ and $\bar{Y}$ are the means of all X_i's and all Y_i's, respectively. OLS will produce different values of $\widehat{\beta_0}$ and $\widehat{\beta_1}$ for different data sets.

4.1.3 Multiple Linear Regression

In many cases, movement in the dependent variable may not be fully explained using only one explanatory variable. We can extend the logic behind single-variable regression analysis to multivariate models by including more explanatory variables into the equation. The theoretical multivariate regression equation has the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i \tag{4.4}$$

where K is the number of explanatory variables. The constant term β_0 indicates the value of the dependent variable when all explanatory variables and the error term equal zero. The coefficients indicate the change in the dependent variable associated with a one-unit increase in the corresponding explanatory variable, holding constant all other explanatory variables in the equation. In other words, multivariate regression enables us to measure the estimated impact of each explanatory variable in isolation.

To illustrate an example, we utilize a sample of 219 college students provided in the *First Year GPA* data set under the *Stat2Data* package in R (Cannon et al., 2018). Some modifications have been made to the data set for our purposes, which are provided in Appendix 3. The variables that will be used to construct the OLS model (as well as other models in the subsequent sections) are listed in Table 4.1:

	Variable Name	**Definition**
Y	GPA	College GPA
X_1	hsGPA	High School GPA
X_2	SATmath	SAT Math Score
X_3	SATverbal	SAT Verbal/Critical Reading Score
X_4	StudyHours	Hours Studied in College per Week
Z	Roommate	Hours Spent with Roommate per Week

Table 4.1: Variables in the College GPA Model

Let our model express first-year college GPA as a linear function of variables listed in Table 4.1: high school GPA, SAT Math score, SAT Verbal and Critical Reading scores, and study hours in college. Note that Z is not included here but will be used as an instrumental variable in the following section. The theoretical equation for this model will have the form:

$$GPA = \beta_0 + \beta_1 \cdot hsGPA + \beta_2 \cdot SATmath + \beta_3 \cdot SATverbal + \beta_4 \cdot StudyHours + \varepsilon \quad (4.5)$$

which can also be understood as the true population regression equation for all first-year college students. Meanwhile, the estimated regression equation will have the form:

$$\widehat{GPA} = \widehat{\beta_0} + \widehat{\beta_1} \cdot hsGPA + \widehat{\beta_2} \cdot SATmath + \widehat{\beta_3} \cdot SATverbal + \widehat{\beta_4} \cdot StudyHours. \quad (4.6)$$

As was stated in Equation 4.3, OLS derives the estimates of the above coefficients in such a way that minimizes the squared residuals. Using a statistical software, we obtain the results in the Table 4.2. Based on the information about the coefficients in Table 4.2, we can now construct the estimated regression equation for the sample of 219 students as follows:

$$\widehat{GPA} = 3.793 + 0.004 \cdot hsGPA - 0.004 \cdot SATmath - 0.0003 \cdot SATverbal + 0.065 \cdot StudyHours. \quad (4.7)$$

Recall that each coefficient indicates the change in the dependent variable associated with a one-unit increase in the corresponding explanatory variable, holding constant all other explanatory variables in the equation. For example,

Table 4.2: College GPA Model (OLS)

	Dependent variable:
	GPA
hsGPA	0.004
	(0.351)
SATmath	-0.004^{*}
	(0.002)
SATverbal	-0.0003
	(0.002)
StudyHours	0.065^{***}
	(0.014)
Constant	3.793^{**}
	(1.611)
Observations	219
R^2	0.122
Adjusted R^2	0.106
Note:	*p<0.1; **p<0.05; ***p<0.01 (Standard errors in parentheses)

an additional hour of study time per week is associated with a 0.065 increase in college GPA, holding the other three variables constant.

4.1.4 Measures of Fit

When presenting regression results, it is common practice to inform the reader of how *well* the model fits the actual data. In most cases, an estimated regression equation cannot fit the data perfectly because real data points are spread out with some degree of randomness. As this randomness cannot be fully captured in a single line, it becomes useful to establish a measure of fit. To construct one, we first introduce a quantity called the **total sum of squares (TSS)**, which is the squared variations of Y around its mean, $\bar{Y}$, and can be computed as follows:

$$TSS = \sum_{i=1}^{N}(Y_i - \bar{Y})^2. \tag{4.8}$$

In Ordinary Least Squares, the TSS can be decomposed into two components. The first component is the **explained sum of squares (ESS)**, which represents the variation in Y that can be explained by the model. ESS can be computed as follows:

$$ESS = \sum_{i=1}^{N}(\widehat{Y_i} - \bar{Y})^2. \tag{4.9}$$

The second component is the **residual sum of squares (RSS)**, which represents the variation in Y that cannot be explained by the model. Recall that residual was defined as the difference between the actual and the estimated values of Y, in other words, $e_i = Y_i - \widehat{Y_i}$. Thus, RSS can be computed as follows:

$$RSS = \sum_{i=1}^{N}(Y_i - \widehat{Y_i})^2 = \sum_{i=1}^{N} e_i^2. \tag{4.10}$$

By definition, total sum of squares is the sum of explained sum of squares and residual sum of squares:

$$TSS = ESS + RSS \tag{4.11}$$

$$\sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \sum_{i=1}^{N}(\widehat{Y_i} - \bar{Y})^2 + \sum_{i=1}^{N}(Y_i - \widehat{Y_i})^2 \tag{4.12}$$

Equation 4.12 states that the total variation in Y can be decomposed into two types of variation: one that can be attributed to the regression equation, and the other that is due to the stochastic nature of the data.

$\mathbf{R}^2$

If the residual sum of squares takes up a large proportion of the total sum of squares, this means that much of the variation in Y cannot be explained by our model. This is equivalent to saying that our estimated regression equation does not fit the data very well. On the other hand, if if the ESS takes up a large proportion of the TSS, this means that our model is capable of explaining much of the variation in Y. This is equivalent to saying that our estimated regression equation fits the data well. The quantity called R^2 measures the overall fit of the model in this manner. R^2 is the ratio of the ESS to the TSS, or equivalently,

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{N} e_i{}^2}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}. \tag{4.13}$$

In Table (2), the R^2 was shown to be 0.122. This means that 12.2 percent of the variation in Y can be explained by our independent variables. This also means that 87.8 percent of the variation cannot be explained by the model due to the incompleteness of the model or simply due to pure randomness within the data.

Adjusted $\mathbf{R}^2$

R^2 may be increased simply owing to the inclusion of more explanatory variables into the model. An additional variable is guaranteed to either increase or have no impact on the explanatory power of the model, but it will never reduce the R^2.

Put differently, adding a variable will either increase the ESS or have no impact on the ESS (if the variable is completely uncorrelated with Y). In empirical applications, increased R^2 is not necessarily a virtue because a nonsensical variable may have been included and increased the ESS. To cope with this problem, a slightly different measure of fit called the **adjusted** R^2 may be used. Adjusted R^2 is derived as follows:

$$Adj.R^2 = 1 - \frac{\sum e_i^2/(N-K-1)}{\sum (Y_i - \bar{Y})^2/(N-1)}. \tag{4.14}$$

The benefit of this measure is that it adjusts for the number of variables in the model, K. In Table 4.2, the R^2 shows to 0.122 but the adjusted R^2 is lower because it accounts for the number of independent variables considered.

4.1.5 Statistical Significance

In essence, estimating a regression equation is equivalent to performing a set of hypothesis tests. In Section X, we created a model of first-year college GPA by considering a set of independent variables. We hypothesized that the variables of our selection would impact our dependent variable, so we used regression analysis to empirically test our hypotheses. Specifically, the hypotheses tested if the coefficient of an independent variable was equal to zero or not. Furthermore, the OLS procedure quantified the level of impact of each explanatory variable on first-year GPA by deriving the coefficients.

As much as we are interested in finding the value of the coefficient estimates, we must pay heed to how much *confidence* we can have in our results. The reliability of our results can be communicated using the concepts of p-value and statistical significance. To define these concepts, let us consider the following hypothesis test for β_1, the coefficient of the *HSGPA* variable:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0.$$

We define **p-value** as the probability of observing something as extreme as or more extreme than what we observed, assuming that H_0 were true. In the case of the GPA model, the p-value of $\widehat{\beta_4}$ is the probability of observing a coefficient estimate for *StudyHours* as extreme as or more extreme than 0.065 , assuming that study hours have no impact on college GPA in actuality. In the meantime, **significance level** α_0 is the probability of rejecting H_0, assuming that H_0 was true. If the significance level is greater than or equal to the the p-value, then we reject H_0. On the other hand, if significance level is less than the p-value, then fail to reject H_0.

As shown in Table 4.2, inequality statements about the p-value of a variable and different significance levels are indicated by different number of asterisks. For example, the *StudyHours* variable has three asterisks, meaning that the p-value is less than or equal to 0.01. This means that chances are less than 1% that we would find *StudyHours* to have an impact as high as or higher than 0.065, assuming that it does not have any impact in actuality. In other words, it is very unlikely that we would have obtained such result just by chance. The p-value being less than or equal to 0.01 allows us conclude that we are 99% confident that *StudyHours* is a statistically significant variable.

4.1.6 The Classical Assumptions

Next, we move on to discuss the **classical assumptions** that are required in order for OLS to be the 'best' estimator for regression analysis. If any of these conditions do not hold true, the validity of our statistical inference may be at risk. Although there exist remedies that can be used when the conditions are not satisfied, it is important to check the conditions by using residual plots. While there exist different statistical tests to check for these assumptions, we will only observe the residual plots in this paper. The classical assumptions are given in Table 4.3:

Table 4.3: The Classical Assumptions

I. The model is linear in the coefficients and has an additive error term.
II. The error term has a population mean of zero.
III. All explanatory variables are uncorrelated with the error term.
IV. Observations of the error term are uncorrelated with each other.
V. The error term has a constant variance.
VI. No explanatory variable is a perfect linear function of another.
VII. The error term is normally distributed.

4.1.7 Residual Plots

A subset of the classical assumptions that specifically concern the error term may be checked using the *plot* function in R. We will examine if these conditions are satisfied in the case of our GPA model. This process will add strength to

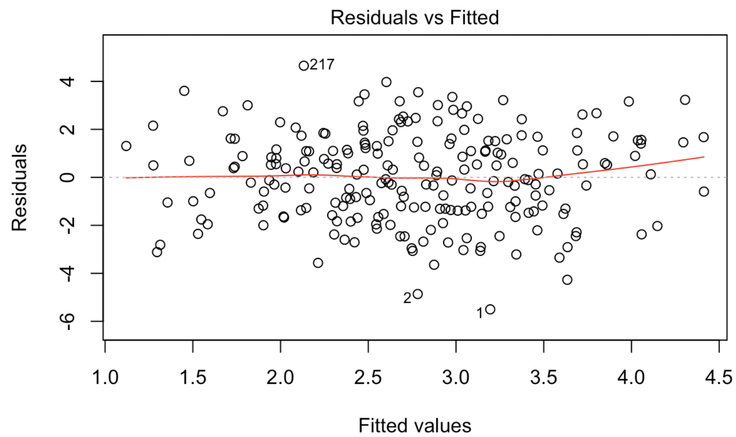

Figure 4.1: Residual Plot for College GPA Model

our selection of OLS as our estimator for linear regression. Since the error term is a theoretical component of the regression model, ε cannot be observed. Alternatively, we can calculate the residuals, e, and perform a residual analysis. While there exists rigorous statistical tests, we will rely on visual presentations in order to communicate the procedure in a more intuitive manner.

Figure B.6 represents the residuals (e_i) of each fitted value ($\widehat{Y_i}$). The red line indicates the calculated mean of the residuals for each fitted value. The first condition, particularly the existence of an additive error term, is met because the error terms are clearly not zero. Also, the second condition is satisfied because the red line closely aligns with the horizontal dotted line where residual equals zero. There is some slight deviance in the both ends of the red line, but we claim this is negligible considering that there are such few points. The fourth condition is also secured as there does not seem to be a correlation between the error terms. Had there been a correlation, the points would have formed a pattern. Lastly, we claim that the fifth condition of error terms having a constant variance is generally fulfilled. Almost all of the data points are distributed

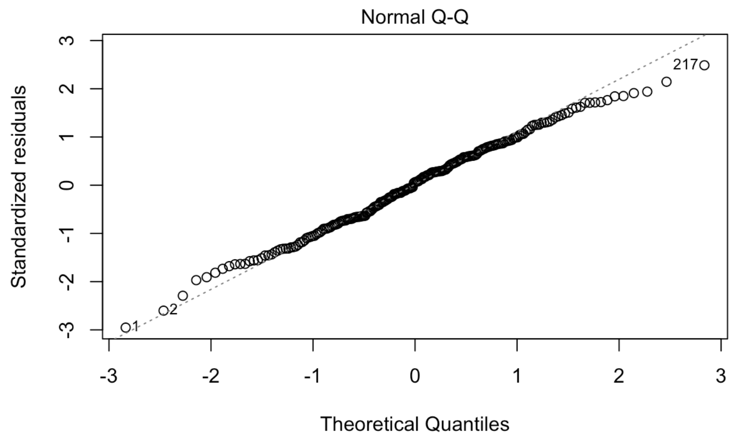

Figure 4.2: Normal Probability Plot

within an imaginary horizontal "band" where residuals are between -4 and 4. The existence of points outside this band certainly weakens the fifth condition, which may affect the statistical significance of our results to a certain extent.

Figure 4.2 enables us to check if the seventh condition is fulfilled. On the y-axis lie the standardized residuals, whereas the theoretical residuals are indicated on the x-axis. Simply put, this plot allows us to compare what actually happened with the residuals versus what should happen if the residuals followed a normal distribution. If the points align with the dotted line, then the normality condition is said to be satisfied. In our case, the points deviate somewhat from the dotted line when theoretical quantiles are greater than 1. Although it is certainly questionable whether the error term "perfectly" follows a normal distribution, we claim that this is not too concerning. The OLS procedure is known to be robust to departures (of the error term) from the normal distribution; simply put, the regression results do not change much even if the seventh condition is not perfectly met (Greene, 2017). This is also confirmed when we draw a histogram of the residuals as in Figure 4.3.

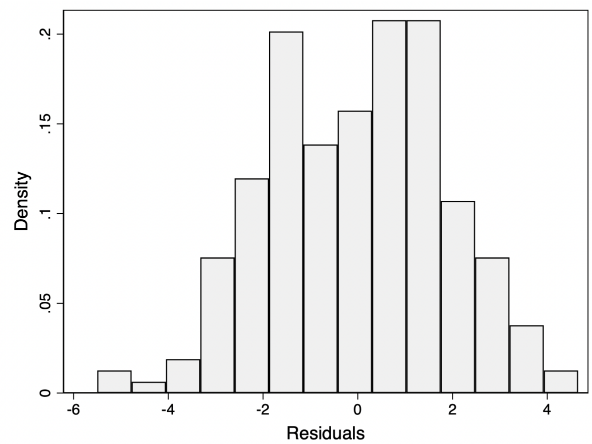

Figure 4.3: Histogram of the Residuals

The distribution is not certainly perfectly normal, however, we claim that it is “close enough”. While it is recommended that the normality condition is checked, minor violations of the condition such as in Figure 4.3 do not drastically reduce the reliability of the regression results (Studenmund, 2017).

4.2 Instrumental Variables

4.2.1 The Endogeneity Condition for OLS

Among the seven conditions required for OLS to be the best estimator, we turn our attention to the third condition which addresses endogeneity. **Endogeneity** refers to the presence of correlation between an explanatory variable and the error term. Recall the theoretical regression equation that was presented in Equation 4.4:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i.$$

An independent variable X_j is said to be **endogenous** if the correlation between X_j and ε_j is non-zero; and **exogenous** if the correlation is zero. If endogeneity is present in the model, OLS produces biased coefficient estimates, which means that the expected value of an estimated coefficient is not equal to the unknown true value of the coefficient. Put differently, endogeneity is problematic because the **bias** in our coefficient estimates becomes non-zero, or equivalently,

$$E(\widehat{\beta_j}) - \beta_j \neq 0.$$

This is not a desirable result since we want our estimates to be as close to the true value of the parameter as possible.

Simultaneous Models

The endogeneity problem commonly occurs in simultaneous models in which variables are jointly determined. To give an example, education level and income may have a two-way causal relationship. On one hand, those with a higher level of education may be more likely to find a higher-paying job. On the other hand, higher income may enable one to afford longer years of education without having to hurriedly get a job. Following Studenmund (2017), we present a generalized form of simultaneous equations that contain such jointly determined variables:

$$Y_{1t} = \alpha_0 + \alpha_1 Y_{2t} + \alpha_2 X_{1t} + \varepsilon_{1t} \tag{4.15}$$

$$Y_{2t} = \beta_0 + \beta_1 Y_{1t} + \beta_2 X_{2t} + \varepsilon_{2t}. \tag{4.16}$$

Equations 4.15 and 4.16 are **structural equations** that are formulated on the basis of theoretical speculation. Similar to the example of education level and income, Y_{1t} and Y_{2t} are jointly determined and are thus endogenous variables. Meanwhile, the X_1 and X_2 variables do not possess this simultaneous nature and are thus exogenous.

Endogeneity in Simultaneous Equations

Simultaneous equations with jointly determined variables violate the endogeneity condition. This violation can be shown rather intuitively through the following train of thoughts. If ε_{1t} increases, then Y_{1t} increases in Equation 4.15; increased Y_{1t} leads to increased Y_{2t} in Equation 4.16; and due to the simultaneous nature, increased Y_{2t} also increases Y_{1t} in Equation 4.15. In short, an increase in ε_{1t} leads to an increase in Y_{2t}, which is an independent variable in Equation 4.15. Due to this correlation between the error term and an explanatory variable, the model expressed in Equation 4.15 is said to violate the endogeneity condition of the OLS. Without loss of generality, the same conclusion can be made about Y_{1t} and ε_{2t}.

4.2.2 2SLS: A remedy for endogeneity

The OLS estimator produces biased coefficient estimates when used in simultaneous models. To cope with this problem, the researcher may choose to use an alternative estimation procedure called **Two-Stage Least Squares (2SLS)**. As the name implies, this method is also based upon the idea of minimizing squared residuals. In essence, 2SLS performs the OLS procedure across two stages while involving the selection of an instrumental variable. To effectively remedy the endogeneity problem, the selected instrumental variable(s) must be *relevant* to the endogenous variable and *exogenous* to the dependent variable. We will examine these conditions in more detail in Section 4.2.3.

Reduced-Form Equations

In Equation 4.15, Y_{2t} is an independent variable that is correlated with the error term, ε_{1t}. To reduce (or, ideally, eliminate) this correlation and resolve the endogeneity problem, we construct another regression equation to express Y_{2t} as a function of some instrumental variables that are uncorrelated with the error term, ε_{1t}. Doing the same for Equation 4.16, we establish a set of **reduced-form equations** that express our endogenous variables, Y_{1t} and Y_{2t}, in terms of the exogenous variables:

$$Y_{1t} = \gamma_0 + \gamma_1 X_{1t} + \gamma_2 X_{2t} + \gamma_3 X_{3t} + \gamma_4 X_{4t} + v_{1t} \tag{4.17}$$

$$Y_{2t} = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t} + \delta_3 X_{3t} + \delta_4 X_{4t} + v_{2t}. \tag{4.18}$$

Note that we include *all* exogenous variables within the system into the reduced-form equations. This is because every exogenous variable in the simultaneous system is a "candidate" to be an instrumental variable. By choosing only one instrumental variable, we would be "throwing away" information (Studenmund, 2017). The γs and δs are the **reduced-form coefficients**, whereas vs are the error terms for the new regression equations. The reduced-form equations do not have inherent simultaneity, so we have successfully avoided the violation of the endogeneity condition. In particular, the variables X_3 and X_4 have been additionally selected as instrumental variables for Y_1 and Y_2, respectively. Now, we proceed to describe the two stages of 2SLS.

2SLS: Stage One

The first step of the 2SLS procedure is to estimate the reduced-form equations using OLS. We estimate Equations 4.17 and 4.18 and write:

$$\widehat{Y_{1t}} = \widehat{\gamma_0} + \widehat{\gamma_1}X_{1t} + \widehat{\gamma_2}X_{2t} + \widehat{\gamma_3}X_{3t} + \widehat{\gamma_4}X_{4t} \tag{4.19}$$

$$\widehat{Y_{2t}} = \widehat{\delta_0} + \widehat{\delta_1}X_{1t} + \widehat{\delta_2}X_{2t} + \widehat{\delta_3}X_{3t} + \widehat{\delta_4}X_{4t}. \tag{4.20}$$

Ideally, the researcher would identify all instrumental variables so that endogeneity is completely eliminated. If this is the case, the X variables would be exogenous and uncorrelated with the error terms, vs. Also, the OLS estimates of the reduced-form coefficients, $\widehat{\delta}$s and $\widehat{\gamma}$s, would be unbiased.

2SLS: Stage Two

In order to perform the second-stage estimation, we must first replace the endogenous variables with the estimated reduced-form equations. That is, we substitute the Ys with the $\widehat{Y}$s (from Equations 4.19 and 4.20) in order to express the structural equations only in terms of the exogenous X variables:

$$Y_{1t} = \alpha_0 + \alpha_1\widehat{Y_{2t}} + \alpha_2X_{1t} + \epsilon_{1t} \tag{4.21}$$

$$Y_{2t} = \beta_0 + \beta_1\widehat{Y_{1t}} + \beta_2X_{2t} + \epsilon_{2t} \tag{4.22}$$

In essence, we are rewriting Equations 4.15 and 4.16 by plugging Equations 4.19 and 4.20 into the $\widehat{Y_{1t}}$ and $\widehat{Y_{2t}}$ terms. As a result, we obtain:

$$Y_{1t} = \alpha'_0 + \alpha'_1 X_{1t} + \alpha'_2 X_{2t} + \alpha'_3 X_{3t} + \alpha'_4 X_{4t} + \epsilon'_{1t} \tag{4.23}$$

$$Y_{2t} = \beta'_0 + \beta'_1 X_{1t} + \beta'_2 X_{2t} + \beta'_3 X_{3t} + \beta'_4 X_{4t} + \epsilon'_{2t} \tag{4.24}$$

where the α' and β' terms represent the coefficients after algebraic simplification. Now, we perform the second-stage of 2SLS by estimating Equations 4.23 and 4.24. Recall that we estimated the first-stage equations with OLS. It must be noted that the second-stage equations are *not* estimated with OLS. Estimating the second-stage equations with OLS will produce incorrect standard errors in our coefficient estimates, $SE(\widehat{\alpha})$ and $SE(\widehat{\beta})$. Therefore, it is important that we use the computer's 2SLS procedure when performing the second stage. In STATA, the built-in command, *ivregress*, performs this procedure.

Properties of 2SLS

There are two properties to note about 2SLS estimation. First, 2SLS estimates are still biased. Simultaneity bias cannot be fully eliminated due to any remaining correlation between the $\widehat{Y}$s produced by the first-stage estimations and the εs. With larger sample size, the 2SLS bias will be reduced but will remain non-zero. However, the expected bias due to 2SLS will be smaller than the expected bias due to OLS. This is certainly an advantage that 2SLS has over OLS. The second property of 2SLS is that the coefficient estimates have increased variances and $SE(\widehat{\beta})$s compared to OLS estimates (Wooldridge, 2016).

4.2.3 Example

We revisit the college GPA example to demonstrate how an instrumental variable can be selected and utilized in a statistical model. Our dependent variable is college GPA with all other independent variables being the same as our previous OLS model described in Equation **??**. One may critique that there exists reverse causality in GPA and study hours. On one hand, students who study more may be likely to have better academic performances. On the other hand, students with high GPA may be more likely to study longer hours to maintain a high GPA. In this sense, one may claim that these two variables are simultaneous in nature. Since *StudyHours* is an endogenous variable, we may consider using an instrumental variable to remedy this problem.

The *roommate* variable was added and simulated into the original data set. This variable represents the number of hours per week spent with a roommate in college. Although not ideal, *roommate* may be an appropriate instrumental variable for two reasons. First, it is theoretically and statistically relevant to the endogenous variable, *StudyHours*. Spending time with a studious roommate may motivate a student to spend more hours studying; having a roommate who loves to play video games may reduce a student's study hours if they spend a lot of time together. When we estimate a simplistic single-variate regression using these variables, we acquire the results in Table 4.4.

The *roommate* variable is statistically significant in the model with an arguably large impact; an additional hour spent with a roommate is associated with a 0.183-point increase in GPA. Thus, we claim that the *relevance* condition is met. Meanwhile, the correlation between *GPA* and *roommate* shows to be -0.06,

Table 4.4: Relevance of *roommate* to *StudyHours*

	Dependent variable:
	StudyHours
roommate	0.183**
	(0.071)
Constant	22.182***
	(1.160)
Observations	219
Adjusted R^2	0.025
Note:	*p<0.1; **p<0.05; ***p<0.01
	(Standard errors in parentheses)

which we consider a very small number for a correlation. Although *roommate* is not purely exogenous to *GPA*, we consider this selection of instrumental variable reasonable and proceed with our example.

As shown in Table 4.5, the coefficient estimate for *StudyHours* changed from 0.065 to 0.125 in the 2SLS model. This change is due to the use of *roommate* as an instrumental variable, which seems to have corrected for the bias was present in the OLS model. As we noted in Section 4.2.2, the standard error of *StudyHours* has increased as a result of the 2SLS procedure.

4.3 Ordered Probit Model

4.3.1 Modeling Binary Outcomes

The Ordinary Least Squares procedure is useful when modelling continuous outcomes. However, at least two problems arise when the dependent variable

Table 4.5: OLS and 2SLS Results

	Dependent variable:	
	GPA	
	OLS	2SLS
hsGPA	0.004	0.197
	(0.351)	(0.475)
SATmath	−0.004*	−0.004*
	(0.002)	(0.002)
SATverbal	−0.0003	−0.0003
	(0.002)	(0.001)
StudyHours	0.065***	0.125*
	(0.014)	(0.084)
Constant	3.793**	1.411
	(1.611)	(3.763)
Observations	219	219
Adjusted R^2	0.106	0.047
Note:		*p<0.1; **p<0.05; ***p<0.01 (Standard errors in parentheses)

involves binary choices or categorical outcomes. To address these problems, we consider the previous GPA model but now with a slightly different dependent variable. What if our data were less complete and did not contain full information about a student's GPA? Let us assume that the dependent variable is a binary indicator of whether a student's GPA is high or low. In other words, our Y equals 0 if a student has a GPA lower than 3.5 and Y equals 1 if their GPA is higher than or equal to 3.5.

The first problem is that our coefficient estimates does not make intuitive sense. If the coefficient of an independent variable, such as high school GPA, is estimated to be 0.05, what does this mean? A 0.05-unit increase in "whether GPA is high or low" may not be as meaningful as OLS coefficients. The second problem is that we may obtain predicted values that are greater than 1 or less than 0. If an individual has all the favorable conditions for a high GPA, then the model may produce a predicted value of 1.2, which is not even a valid outcome.

To cope with the first problem, we may look at the above linear regression model as a **linear probability model**. In other words, we treat the binary dependent variable as the *probability* of being employed. This allows us to interpret a coefficient estimate of 0.05 as a 5% increase in the probability of being employed. However, the second problem remains. Two of the most commonly adopted solutions to this problem are the **logit** and **probit** models.

4.3.2 Different Link Functions

The strongest motivation for using a logit or a probit model would be to constrain the dependent variable within the range between 0 and 1. To achieve this result, the logit model uses the link function:

$$\text{Prob}(Y = 1|\mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{\exp(\mathbf{x}'\boldsymbol{\beta}) + 1} \tag{4.25}$$

where $\boldsymbol{\beta}$ is the set of parameters that show the impact of changes in independent variables $\mathbf{x}$ on the probability. For the sake of concision, we express our explanatory variables as the vector $\mathbf{x}$, instead of listing them out individually.

Meanwhile, the probit model uses the link function:

$$\text{Prob}(Y = 1|\mathbf{x}) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \phi(t)dt = \Phi(\mathbf{x}'\boldsymbol{\beta}) \tag{4.26}$$

where ϕ and Φ are the probability distribution function (pdf) and cumulative distribution function (cdf) of the normal distribution, respectively. Greene (2016) and Studenmend (2017) note that the partial effects produced by these two models are nearly the same. In the vein of Jaldell (2005; 2019), we will proceed with the probit model only.

4.3.3 Binomial Probit Model

We begin with the most basic form of the probit model, which involves a binary independent variable. We will discuss the structure of the binomial probit model and extend our consideration to the ordered probit model which will have multiple levels within the dependent variable.

The Latent Variable

A **latent variable** is a variable "behind the scenes" that determines the value endowed to the dependent variable. In many cases, the latent variable is not observable. When modelling the *choice* that an individual makes between two options, the latent variable is the net utility of her decision. The individual is assumed to gauge the cost and benefit of each option, and choose the option that produces a net benefit greater than 0. On the other hand, when modelling two possible *outcomes*, the latent variable would be an unobservable measure of the outcome. For example, to model the severity of a fire outcome, some theoretical measure of the severity of fire spread will be implicitly assumed to be the latent variable. If the severity of a fire outcome exceeds a certain threshold, then a value of 1 (or 0) will be endowed to the dependent variable Y.

The Latent Regression Model

We now consider an underlying latent regression equation. Let the latent variable y^* be a function of $\mathbf{x}$ and $\boldsymbol{\beta}$, each being a matrix of independent variables and their corresponding coefficients, respectively. In other words,

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

where we assume that ε follows a standard normal distribution with mean zero and variance one. The right-hand side of the above equation is equivalent to that of Equation 4.4. Recall that we do not observe the actual values of y^* but only whether a choice is made or not (for binary choice models) or whether an outcome fits into a certain category or not (for binary outcome models). If the net utility of a choice is greater than zero, then a value of 1 is endowed to the dependent variable of the (actual, not underlying) model to indicate that the choice is made. If the net utility is less than or equal to zero, then we endow a 0 to our Y.

Likewise, in a binary outcome model, we endow a 1 or 0 to our dependent variable, y, depending on whether the latent variable, y^*, is above or below a certain threshold. To simplify our discussion, we may assume this threshold to be 0.

Functional Form and Probability

Now, let us examine how Equation 4.26 can be obtained. Consider that we have defined our dependent variable so that:

$$Y = 1 \text{ if } y^* > 0,$$

$$Y = 0 \text{ if } y^* \leq 0.$$

To solve for the first case where $y^* > 0$, we write:

$$y^* = \mathbf{x'}\boldsymbol{\beta} + \varepsilon > 0.$$

Rearranging and solving for ε, we obtain:

$$\varepsilon > -\mathbf{x'}\boldsymbol{\beta}.$$

Since we assumed ε to follow a normal distribution, we express the *probability* of $Y = 1$ in terms of the area underneath the normal distribution curve, where the threshold is 0. As a result, we obtain the probability statement:

$$\text{Prob}(Y = 1|\mathbf{x}) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(t)dt = \Phi(\mathbf{x'}\boldsymbol{\beta}).$$

By a similar logic, we can express the probability of $Y = 0$ as follows:

$$\text{Prob}(Y = 0|\mathbf{x}) = \int_{\mathbf{x}'\beta}^{\infty} \phi(t)dt = 1 - \Phi(\mathbf{x'}\boldsymbol{\beta}).$$

Recall that predicted values in a binary outcome model represent the probability of an event. We have shown that the link function for the probit model can be derived from the underlying definition of our dependent variable Y. Thus, we conclude with the following:

$$\text{Prob}(Y = 1|\mathbf{x}) = \Phi(\mathbf{x}'\boldsymbol{\beta}) \tag{4.27}$$

$$\text{Prob}(Y = 0|\mathbf{x}) = 1 - \Phi(\mathbf{x}'\boldsymbol{\beta}), \tag{4.28}$$

where $\boldsymbol{\beta}$ is the set of coefficients that indicate the impact of changes in $\mathbf{x}$ on the probability.

Estimation of Beta Coefficients

The probit model cannot be estimated under the ordinary least squares procedure because our dependent variable is not continuous. Essentially, our data points do not really have a y-coordinate since the dependent variable is a binary variable. Therefore, the concept of minimizing the distance to a line of best fit is not relevant in the first place. Instead, we use the maximum likelihood estimator to derive the coefficient estimates. As the name implies, the objective of maximum likelihood estimation is to maximize the likelihood function so that our observed data is the most probable.

We can construct the likelihood function by considering each choice or outcome as a Bernoulli trial. We use the probability statements that were presented in the previous section.

Let y_i represent each of the Bernouilli outcomes for the i^{th} observation. In a binary case, y_i is either 0 or 1. Also, using the the probability statements we have established, we can express the likelihood of the i^{th} observation as follows:

$$\mathrm{Prob}(Y = y_i|\mathbf{x}) = [\Phi(\mathbf{x}'\boldsymbol{\beta})]^{y_i} \cdot [\Phi(1 - \mathbf{x}'\boldsymbol{\beta})]^{1-y_i}.$$

As a result, we construct the likelihood function as follows:

$$L = \prod_{i=1}^{N} [\Phi(\mathbf{x}'\boldsymbol{\beta})]^{y_i} \cdot [\Phi(1 - \mathbf{x}'\boldsymbol{\beta})]^{1-y_i} \tag{4.29}$$

where N represents the number of observations, or equivalently, Bernouilli trials. The maximum likelihood estimator can be obtained by maximizing the log of the likelihood function L. Since there is no closed form solution to this optimization problem, a numerical approximation may be used.

Calculating Marginal Effects

Just as any other regression analyses, the goal of the probit model is to quantify the impact of an independent variable on the outcome variable. Since the coefficient estimates do not convey this information, marginal (partial) effects are calculated in many applications. By definition, marginal effect is the partial derivative of the cumulative distribution function with respect to an independent variable, where the dependent variable Y takes a particular value. The vector of marginal effects can be obtained using the chain rule as follows:

$$\frac{\partial \Phi(y|\mathbf{x})}{\partial \mathbf{x}} = \Big[\frac{\partial \Phi(\mathbf{x}'\boldsymbol{\beta})}{\partial \mathbf{x}}\Big] \times \boldsymbol{\beta} = \phi(\mathbf{x}'\boldsymbol{\beta}) \times \boldsymbol{\beta} \tag{4.30}$$

where Φ and ϕ are the cdf and pdf of the normal distribution, respectively.

There are largely two ways of computing partial effects. The first is the partial effects at the averages (PEA), for which the partial derivatives are evaluated at the the sample means of the independent variables. In other words,

$$PEA = \phi(\bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}) \cdot \hat{\boldsymbol{\beta}}.$$

PEA becomes less meaningful when the independent variables are binary variables. For instance, there is no individual in the sample with a "mean gender" of 0.5. For this reason, the average partial effects (APE) are more commonly used, which are calculated as follows:

$$APE = \frac{1}{n} \sum_{i=1}^{n} \phi(\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}) \cdot \hat{\boldsymbol{\beta}}.$$

Simply put, APE is the average of the partial effects for all observations in the sample.

4.3.4 Ordered Probit Model

The binary probit model can be extended to consider the cases in which there are multiple levels within the dependent variable. A necessary condition for this extension is the proportional odds assumption, or the parallel regression assumption. Ordered probit models are built upon the assumption that the relationship between any pairs of outcome categories is the "same". However, as Greene (2018) points out, the difference between $Y = 0$ and $Y = 1$ versus $Y = 1$ and $Y = 2$ may not necessarily be the same. In many cases, these numbers often function as a categorical distinction for the different levels of outcomes rather than containing a numerical meaning. Therefore, the researcher must be aware of the proportional odds assumption and the potential limitations of her chosen method.

The simplest form of an ordered probit model would be the one with three levels. To be divided into three levels or categories, two cutoff values are needed. For the sake of simplicity, we may assume the first cutoff to be zero, so that only one unknown threshold constant, μ, is needed.

Consider the unobservable latent variable, y^*. The associated y values are as follows:

$$Y = 0 \text{ if } y^* \leq 0,$$
$$Y = 1 \text{ if } 0 < y^* \leq \mu,$$
$$Y = 2 \text{ if } \mu \leq y^*.$$

Using the fact that $y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$ and that ε follows a normal distribution, we can construct the functional form of the probit model as follows:

$$\text{Prob}(Y = 0|\mathbf{x}) = 1 - \Phi(\mathbf{x}'\boldsymbol{\beta}) \tag{4.31}$$
$$\text{Prob}(Y = 1|\mathbf{x}) = \Phi(\mu - \mathbf{x}'\boldsymbol{\beta}) - \Phi(-\mathbf{x}'\boldsymbol{\beta}) \tag{4.32}$$
$$\text{Prob}(Y = 2|\mathbf{x}) = 1 - \Phi(\mu - \mathbf{x}'\boldsymbol{\beta}). \tag{4.33}$$

For these probabilities, the marginal effects of changes in the independent variables can be calculated in a similar fashion from Equation 4.30:

$$\frac{\partial \Phi(Y = 0|\mathbf{x})}{\partial \mathbf{x}} = -\phi(\mathbf{x}'\boldsymbol{\beta}) \cdot \boldsymbol{\beta}, \tag{4.34}$$
$$\frac{\partial Phi(Y = 1|\mathbf{x})}{\partial \mathbf{x}} = \left[\phi(-\mathbf{x}'\boldsymbol{\beta}) - \phi(\mu - \mathbf{x}'\boldsymbol{\beta})\right] \cdot \boldsymbol{\beta}, \tag{4.35}$$
$$\frac{\partial Phi(Y = 2|\mathbf{x})}{\partial \mathbf{x}} = \phi(\mu - \mathbf{x}'\boldsymbol{\beta}) \cdot \boldsymbol{\beta}. \tag{4.36}$$

4.3.5 Example

To demonstrate an example, we will continue to use the *First Year GPA* data set. As was mentioned in Section 4.3.1, we will construct a binary variable named *HighGPA* and use this as our dependent variable. The variable will equal 0 if a student's GPA is lower than 3.5 and 1 if it is higher than or equal to 3.5. As the coefficient estimates cannot be interpreted as marginal effects, we will use the *margins* command in STATA to produce the partial effects, specifically the average partial effects.

Table 4.6: Probit Model Results

	Dependent variable:	
	GPA: High or Low	
	Coefficients	Partial Effects
hsGPA	1.051***	0.288***
	(0.306)	(0.080)
SATmath	−0.0002	−0.00005
	(0.002)	(0.004)
SATverbal	0.003	0.0006
	(0.002)	(0.0004)
StudyHours	−0.021*	−0.006*
	(0.011)	(0.003)
Constant	3.793***	-
	(1.390)	-
Observations	219	219
Pseudo R^2	0.110	0.110
Note:	*p<0.1; **p<0.05; ***p<0.01 (Standard errors in parentheses)	

The results in Table 4.6 show that a one-unit increase in high school GPA is associated with a 28 percentage point increase in the *probability* of having a high college GPA (in other words, a GPA of 3.5 or higher). As probit models are not estimated with OLS, the pseudo R^2 does *not* represent the proportion of variance of the independent variable explained by the explanatory variables. Each statistical software package may report a different type of pseudo R^2

established by different statisticians. Greene (2017) provides further detail on this matter.

4.4 Application to Current Study

In this chapter, we have carefully examined three statistical methods: multiple linear regression using Ordinary Least Squares, Two-Stage Least Squares estimation using an instrumental variable, and the ordered probit model. These methods were chosen because of their potential relevance to the subject matter. Multiple linear regression will be used to describe the composition of intermediate outputs in the case of fire service provision. One benefit of this method is the intuitive interpretability of coefficient estimates, which may shed light on the relationship between the inputs and outputs. In the meantime, a set of ordered levels had to be considered as the dependent variable due to a lack of detailed information about fire spread. Since the dependent variable is not continuous in nature, the ordered probit model will prove useful in creating a production function for fire service output.

The instrumental variable method emerged as a promising candidate for this study but ended up not being used at all. The initial anticipation was that the 2SLS method may help mitigate the endogeneity of the fire truck variable (which indicates the number of fire trucks involved in a fire incident). However, while it is reasonable to expect a larger number of fire trucks to produce a higher level of service output, it is also true that a larger number of fire trucks implies that fire was already very severe to begin with. Due to limited data on the fire truck variable, it was difficult to resolve the endogeneity problem using the instrumental variable method.

Having established a firm understanding of the methodology, we now proceed to specify our empirical models in the following chapter.

Chapter 5

Models and Data

This chapter presents the model specification and the data used to estimate the models.

5.1 Model Specification

In Chapter 2, we established firefighters, fire trucks, response time, and suppression time as intermediate *outputs*, all of which become factor *inputs* for producing service output. This provided a theoretical justification for including the intermediate outputs as independent variables within the service output model.

In an attempt to illuminate the relationship between variables, we will first specify the models for all **intermediate outputs**: dispatch level, response time, and suppression time. In all of these models, a linear relationship is assumed between the dependent variable and the independent variables. Note that this paper focuses only on the *ex post* output rather than the *ex ante*. In other words, the stand-by level, G, will not be considered in this chapter as it is the *ex ante* first-stage output. The stand-by level consists of firefighters and fire trucks distributed to each city *before* the incidence of a fire. Specifying a model for G necessarily involves the 'prevention' aspect of the fire services, which is beyond the scope of this paper.

Subsequently, a model for **service output** will be specified. While this model largely follows the work of Jaldell (2005), recall that two differences were noted in Chapter 3. First, we suggested an alternative definition of the starting point of fire when constructing the dependent variable. That is, rather than considering the location of fire upon a fire service's arrival, we claimed that considering the location of fire when the emergency call was made would be more appropriate. Second, we pointed out that municipality fixed effects are likely to absorb not only each fire service's productivity level but also many other aspects of the municipality, including institutional characteristics.

5.1.1 Dispatch Level

In Equation 2.3, dispatch level was defined as the subset of firefighters, fire trucks, and equipment dispatched to the fire scene:

$$G' = g(L', K', Z').$$

The dispatch level in a given fire incident has two components: the **initial crew**, G'_i, and the **additional crew**, G'_a, who back up the initial crew upon their request.

We may express this composition as:

$$G' = G'_i + G'_a, \tag{5.1}$$

which allows us to consider the different sets of factors and decisions that affect G'_i and G'_a, respectively. We present the theoretical regression equation for dispatch level, G', as follows:

$$\begin{aligned} \mathit{DispatchLevel} = \beta_0 &+ \beta_1 \cdot \mathit{Building} + \beta_2 \cdot \mathit{Hazard} + \beta_3 \cdot \mathit{Nearby} \\ &+ \beta_4 \cdot \mathit{SameCalls} + \beta_5 \cdot \mathit{ReportedSeverity} \\ &+ \beta_6 \cdot \mathit{Urgency} + \beta_7 \cdot \mathit{ActualSeverity} + \varepsilon. \end{aligned} \tag{5.2}$$

The first six independent variables affect G'_i and the last affects G'_a. The definitions of the variables are provided in Table 5.1.

Variable	**Definition**
Building	Type, structure, and height of building on fire
Hazard	Presence of hazardous or flammable materials
Nearby	Number of fire stations within 3km radius
ReportedSeverity	Severity or size of fire described by the caller
SameCalls	Number of emergency calls for the same fire
Urgency	Degree of urgency in caller's description
ActualSeverity	Severity or size of fire described by firefighters

Table 5.1: Dispatch Level Model: Variable Definitions

The decision relating to the level of initial crew is made by the **emergency call dispatcher** who picks up the call, listens to the description of the fire, and sends out certain types and number of fire trucks. The *types* of fire vehicles and equipment to be dispatched are mostly determined by the described structure or height of the building. For example, if the fire is said to be on the tenth floor, a ladder truck will be included in the initial dispatch level.

The *number* of fire vehicles can be dependent upon the potentially subjective decisions of the dispatcher. If the fire is described to involve hazardous materials or to be producing a large amount of smoke, a larger number of fire trucks may be dispatched. Also, if the reporter of the fire (who made the emergency call) seems to be in great horror or urgency, the dispatcher may be inclined to perceive the fire as more severe and send out more trucks. According to fire experts, the number of redundant emergency calls also affects the dispatch level. The explanation is that if there are multiple individuals reporting the same fire, the dispatcher anticipates the fire to be very severe and sends out more trucks. In addition, experts explain that dispatching an additional unit of fire service becomes "less burdensome" for the dispatcher if there exists another fire station in close proximity to the primary unit (that is responding to the fire in its own jurisdictional unit). Since the 'cost' of sending out an additional fire service is low, the dispatcher may rather send out 'too many' trucks than send out too few and receive criticism for misjudgement.

In the meantime, the number of additional crew, G'_a, is largely determined by the members of the **initial crew**. The firefighters at the scene may request for back-ups if the fire is too large or severe. The dispatcher almost fully relies on the radio messages from the initial crew when sending additional fire trucks.

The error term in Equation 5.2 may include individual differences between dispatchers, such as risk-averse characteristics or level of experience; productivity differences between one initial crew and another; and other stochastic elements.

Due to insufficient data, the model for dispatch level could not be estimated. In fact, some of the independent variables are not observable or measurable because they are involved with the subjective perceptions or judgements of the dispatcher. Nevertheless, we have specified this model because it will prove useful for identifying the sources of endogeneity when we consider the service output model in later sections.

5.1.2 Response Time

Response time refers to the time it takes the first fire service to arrive at fire scene after the emergency call is made. In Section 2.5, the model for estimated response time (Kolesar, 1975) was introduced as a useful tool with which decision makers can determine the number and location of fire stations to be established across a city. In this section, we consider an *ex post* framework and model the *actual* response times rather than expected response times.

The theoretical regression equation for the response time model can be written as follows:

$$\begin{aligned} RespTime = \beta_0 &+ \beta_1 \cdot Distance + \beta_2 \cdot StationPopDen + \beta_3 \cdot Highway \\ &+ \beta_4 \cdot RoadQuality + \beta_5 \cdot RushHour + \beta_6 \cdot WeekDay \\ &+ \beta_7 \cdot Campaigns + \beta_8 \cdot BuildingType + \beta_9 \cdot FireType \\ &+ \beta_{10} \cdot Weather + \varepsilon. \end{aligned} \tag{5.3}$$

The definitions of the independent variables are given in Table 5.1.2.

The **distance** from the fire station to the fire scene is included as the first

Variable	Definition
Distance	Distance from fire station to fire scene (km)
StationPopDen	Mean population in 1km^2 covered by a city's fire station
Highway	Proportion of distance travelled on a highway
RoadQuality	Condition of roads such as age or width
RushHour	Whether fire occurred during rush hour
WeekDay	Day of week
Campaigns	Number of monthly road campaigns by fire department
BuildingType	Type of building (residential, public, industrial, etc.)
FireType	Type of fire (explosive, chemical, smoke alarm, etc.)
Weather	Weather conditions (wind, rain, and snow)

Table 5.2: Response Time Model: Variable Definition

explanatory variable. As established in Section 2.3.2, this distance exists as a product of the spatial allocation decisions that are made prior to the fire. If this allocation decision was optimal, then distance would generally be shorter because the fire stations would have been placed in greater proximity to sub-regions with higher fire risk. All other things being equal, longer distance is expected to increase response time.

The *StationPopDen* of each city fire department refers to the average population density (persons/km^2) per fire station, or equivalently:

$$StationPopDen = \frac{population}{area} \times \frac{1}{stations} \times \frac{1}{100}.$$

We multiplied by $\frac{1}{100}$ simply for calculation reasons. We expect that having more fire stations concentrated in a given area will decrease the response time. Note that 'population density' was considered inadequate for this model because it does not contain any information on how many fire stations are available in a city. For this reason, we *divided* population density by the number of fire stations.

Even if the distance is the same, the travel time will differ according to **road conditions**. For example, if the fire trucks travel on a highway, the response time may be shorter even if the distance is longer. Also, if the roads are generally older and narrower, then this will slow down the speed at which firefighters can travel. For these reasons, the *Highway* and *RoadQuality* variables are included in the theoretical equation. However, due to a lack of data, these two could not be included in our estimation. In the meantime, certain weather conditions may prolong response time. Indeed, safety is considered as crucial as promptness when it comes to the delivery of fire service. Thus, strong wind or extreme snow may cause the fire trucks to travel at a lower speed to avoid accidents.

Similarly, **traffic conditions** may affect response time. Travel speed may be reduced during rush hours or on busier days of the week. In addition, how quickly citizen drivers yield the right-of-way to fire trucks could be a factor. In South Korea, citizens are not legally obligated to pull over to the right or stop parallel to the curb. Thus, city fire departments conduct monthly campaigns to educate or remind citizens of their social responsibilities. Depending on the number and effectiveness of such campaigns, the behavioral responsiveness of citizen drivers to emergency vehicles may vary, which might also affect response time. Unfortunately, the *Campaign* variable could not be included in the estimation as data were unavailable.

While firefighters are trained to respond to all fires in an equally prompt manner, it is still possible that the degree of urgency with which they travel to the fire scene may be affected by the description of the **fire** or type of the **building** on fire. For example, an identical group of firefighters responding to a smoke alarm and to a massive factory fire may experience different levels of tension until they arrive at the scene, which may lead to varied response times. For this reason, the *BuildingType* and *FireType* variables are considered in the model.

5.1.3 Suppression Time

The theoretical regression equation for the suppression time model can be written as follows:

$$\begin{aligned} SuppTime = \beta_0 &+ \beta_1 \cdot StartPoint + \beta_2 \cdot RespTime + \beta_3 \cdot DispatchLevel \\ &+ \beta_4 \cdot OtherServices + \beta_5 \cdot Hydrant + \beta_6 \cdot Hazard \\ &+ \beta_7 \cdot Weather + \beta_8 \cdot RescueNeed + \beta_9 \cdot Teamwork + \varepsilon \end{aligned} \tag{5.4}$$

where the definition of each independent variable is given in Table 5.3:

Suppression time, measured in minutes, will largely be affected by the **initial severity** of a fire. Since there does not exist an objective, observable 'measure' of fire severity, information about the starting point of a fire may be considered. The variable *StartPoint* indicates the location of fire within the house when the emergency call is made, further detail for which will be given in Section 5.2. If a fire is more severe to begin with, it may take longer to suppress the fire. For example, if the fire is only contained in the room, then the suppression will be relatively easy compared to a fire that had already spread out to the entire building when the emergency call was being made.

Variable	Definition
StartPoint	Location of fire at the time of emergency call
RespTime	Response time in minutes
DispatchLevel	Number of firefighters and fire vehicles at fire scene
OtherServices	Workers and equipment from other public services
StationDensity	Mean area covered by a city's fire station (km^2)
Hydrant	Number of hydrants per km^2 across the city
Hazard	Presence of explosive or inflammable substances
Weather	Weather conditions (temperature, humidity, wind speed)
RescueNeed	Number of persons who needed to be rescued
Teamwork	Teamwork among firefighters dispatched to a scene

Table 5.3: Suppression Time Model: Variable Definitions

Components of the *ex post* firefighting power will also be factors of suppression time. On one hand, prolonged **response time** will allow the fire to spread more, which may require more suppression work and result in increased suppression time. On the other hand, a greater number of firefighters and firetrucks will increase the firefighting power of a fire service, meaning that the suppression activity may be more effective. In other words, greater **dispatch level** is expected to decrease suppression time. At a fire scene, having one more crew member or one more vehicle can make a big difference to the ease of firefighting activity. For example, an additional person may enhance the mobility of firefighters holding the water hoses, as a hose can weigh up to a thousand pounds depending on water pressure. Also, the presence of a tank truck connected to a pump truck provides a direct source of water, without firefighters having to look for a hydrant nearby. Volunteer firefighters[1] present at the fire scene were

[1] In South Korea, each city has a volunteer fire service unit consisting of volunteer, non-

also included for a similar reason.

Pumper and tank trucks are the most prevalent types of fire engines, equipped in almost every fire station. In addition to these, a city fire department also has the following vehicles in place: ladder truck, chemical truck, rescue truck, and ambulance. The provincial headquarters also sends out a helicopter when needed. These vehicles are all included in the *DispatchLevel* variable. In particular, the rescue and paramedic squads were included because they can indirectly enhance the firefighting power by lessening the burden on firefighters. Due to their presence, firefighters can focus on fire suppression and worry less about the rescue or paramedic aspects.

In the meantime, *OtherServices* include all of the following types of workers and equipment from other public services: the office of electricity and gas, the police, soldiers, municipal offices (such as the city hall), and 'others'. These personnel and their equipment may also contribute to enhanced firefighting power by enabling firefighters to concentrate their effort into fire suppression.

The *StationDensity* variable indicates the average land size (km^2) that each fire station of a particular city covers, or equivalently:

$$StationDensity = \frac{area}{stations}.$$

This variable contains information about how promptly an additional unit of fire service can arrive at the scene to back up the primary unit. Since our data set does not contain this information, *StationDensity* was created as a proxy. Notice that the formula for this variable is the very definition of *service reach* that we

professional, citizen firefighters are who are often dispatched to the scene.

introduced in Section 2.5. The time it takes additional units of firefighters to arrive at the scene will be shortened if the average service reach of a fire station is smaller. The *StationDensity* variable captures this density of fire stations across a city. We hypothesize that a higher density of fire stations will produce greater firefighting power and result in shorter suppression time.

External factors are also considered in the model. Extreme weather conditions such as high **temperature** or high **humidity** are likely to diminish the labor productivity of firefighters. In the summer, the protective gears trap a great deal of external heat and can cause the firefighters to experience a temperature as high as 50°C (122 °F). High humidity may also negatively impact respiration and reduce the effectiveness of the firefighting activity. High **wind speed** may be of a hindrance to effective suppression as fire could spread in unexpected directions and aggravate the situation.

The presence of **hazardous materials** may prolong the suppression time in two ways. First, the firefighters may have to be more cautious and strategic about their approach, which can slow down the firefighting process. Second, explosive objects or inflammable substances may simply increase the severity of a fire, which will then increase the time it takes to put out the fire.

The availability of **fire hydrants** may be a factor that reduces suppression time. When suppressing severe fires, pumper trucks may run out of water. If hydrants are located in many places across the city, replenishment of water can be made more quickly without having to wait for additional tank trucks. As a result, suppression time may be reduced.

Also, **teamwork** can play a crucial role in the effectiveness of suppression activity. At a fire scene, collaboration may occur not only *within* a fire station but also *across* fire stations. Since *all* fire services in our sample are required to hold the same number of monthly fire drills and training sessions, no variation is captured across different fire departments. Also, it is difficult to observe or measure teamwork among firefighters from different stations. Thus, the *Teamwork* variable could not be operationalized.

Lastly, the presence of one or more persons waiting to be **rescued** may negatively impact suppression time (Jaldell, 2005). Firefighters will prioritize life-saving over fire suppression when a person is inside a structure that is on fire. For this reason, the number of persons requiring rescue activity was included in the model.

5.1.4 Service Output

Recall that service output, Q, is the final form of output that we have defined in our theoretical framework. In particular, we established labor, capital, other materials, response time, and suppression time as the five factors of production, which was expressed in Equation 2.5:

$$Q = q(L', K', Z', R, T).$$

While these inputs will necessarily be involved in the production of service output, different sets of control variables should be considered for different types of fires. For example, the set of factors that affect service output in residential fires will certainly be different from that of industrial fires. Therefore, service output should be considered according to the type of building or object on fire. Given the types of fires available in our data set, the composition of service output, Q, may be written as follows:

$$
\begin{aligned}
Q = Q_{residential} + Q_{industrial} + Q_{public} + Q_{vehicle} \\
+ Q_{commercial} + \cdots + Q_{mountain} + Q_{other}.
\end{aligned}
\tag{5.5}
$$

Given the scope of this paper, we will only consider residential fires and aim to establish a production function for $Q_{residential}$.

Finally, we present the theoretical regression equation for service output in the case of residential fires:

$$
y_{it} = \beta_0 + \sum_{m=1}^{8}(\beta_m \cdot x_{mit}) + \sum_{n=1}^{16}(\gamma_n \cdot z_{nit}) + \alpha_i + \delta_i + \varepsilon_{it} \tag{5.6}
$$

where y_{it} is the service output that is produced in the t^{th} fire incident of the i^{th} city fire department; x_m are the key independent variables that represent the factors of production; z_n includes all exogenous control variables; and α represents the average service reach of fire stations in a city. The names and definitions of the variables are presented in Table 5.4.

While the x_m variables represent the factors of production established in our theoretical framework, labor inputs are *not* included in the model. When the emergency call dispatcher sends out one or more fire services to the fire scene, the dispatcher does not really consider the number of firefighters, per se; each firefighter is designated to a particular fire vehicle anyway, so the dispatcher bases their decision on the **fire trucks**. For example, in South Korea, three to four firefighters are designated to board a pumper truck when the vehicle is dispatched to an incident, while two firefighters board on each tank truck.

	Variable	Definition
y_{it}	*ServiceOutput*	Service output produced in t^{th} fire dispatch of city i
x_2	*Pumper*	Total number of pumper trucks (mobility)
x_2	*PumperSq*	Total number of pumper trucks squared
x_3	*Tank*	Total number of tank trucks (water source)
x_3	*TankSq*	Total number of tank trucks squared
x_4	*Ladder*	Total number of ladder trucks (height)
x_5	*RespTime*	Response time in minutes
x_6	*SuppTime*	Suppression time in minutes
x_7	*Hydrant*	Number of fire hydrants per km^2
x_8	*OtherInputs*	Personnel and vehicles from other public services)
z_1	*Floor*	Floor on which the fire started
z_2	*RushHour*	Whether fire occurred during rush hours
z_3	*WeekDay*	Day of week
z_4	*Temperature*	Temperature (degrees in Celsius)
z_5	*Humidity*	Humidity (grams of water vapor per kilogram of air)
z_6	*WindSpeed*	Wind speed (meters per second)
z_7	*Snow*	Weather alert for extreme snow
z_8	*Rain*	Weather alert for extreme rain
z_9	*RescueNeed*	Number of persons who needed to be rescued
z_{10}	*Prevention*	Whether equipped with prevention facilities
z_{11}	*Kitchen*	Whether fire started at kitchen
z_{12}	*Trash*	Whether fire started on trash
z_{13}	*Electricity*	Whether fire started on electric object or circuit
z_{14}	*Gas*	Whether fire involved gas leakage or toxic gas
z_{15}	*Arson*	Whether fire was due to arson
z_{16}	*Explosive*	Presence of explosive material
α_i	*StationDensity*	Mean area covered by a fire station in city i (km^2)
δ_i	*PopDensity*	Population density of city i (persons/km^2)

Table 5.4: Service Output Model: Variable Definition

Since *Firefighter* is a linear combination of all the (vectors of) truck variables, it was considered more reasonable to exclude *Firefighter* from the model. A relatively high correlation of 0.75 between *Firefighter* and *Pumper* supports this exclusion.

Response time and **suppression time** are both considered in the service output model. In previous chapters, we have established these two time factors as factors of production.

The *Pumper* and *Tank* variables have their square terms included in the model. The purpose of squaring is to reflect the potential **congestion** issues that may arise when too many fire trucks congregate at the fire scene. This may be a reasonable consideration particularly for urban areas with higher population density. In Equation 5.2, we expected a positive relationship between *DispatchLevel* and *Nearby*; that is, the dispatch level would be higher if an additional unit of fire service is located close to the primary unit. As a result, it is often the case that a larger number of fire trucks tend to be dispatched in densely populated areas. If there are too many fire trucks within a limited space, then the marginal productivity of capital may decrease due to congestion. Since there are typically only one or two ladder trucks within a city, the *Ladder* variable was not squared.

Most components of 'other inputs', Z, are equipment and materials that are contained or built into the fire trucks. While these are not included in the model, we did consider the number of hydrants per km^2 in each city as the Z variable. Meanwhile, the *OtherInputs* variable includes all other personnel and vehicles from the suppression time model: additional emergency vehicles equipped in the main fire department building, helicopter dispatched from the provincial headquarters, civil servants from other public services, and volunteer firefighters.

The z_m variables include a comprehensive list of control variables that are expected to affect service output in residential fires. If a fire occurs on a high level of **floor** within a building, fire suppression will become more challenging. This may lead to worse outcomes and result in lower level of service output[2].

External factors such as **weather** and **traffic** and the *RescueNeed* variable are included in the service output model for the same reasons as in the previous model. As firefighters value life-saving over protecting property, a need for **rescue** activity may negatively impact service output. This statement may sound counter-intuitive because saving a life is a desirable outcome. However, our dependent variable measures the severity of fire spread not the number of lives saved. This will be further elaborated in Section 5.3 when we construct the dependent variable.

Some residential buildings are subject to regulations that require **prevention** facilities or devices within the building, such as sprinklers or alarm systems. The binary variable *Prevention* indicates whether such regulations apply. The variable is expected to have a positive relationship with service output because prevention facilities help mitigate the damage of a fire and may lead to a more desirable outcome.

Various causes and sources of fire were considered, as represented in variables z_{11} through z_{16}. We anticipate that the impact of miscellaneous fires, such as trash or food on the stove, would be less critical than that of fires which

[2] The dependent variable not only measures the *amount* of work performed but also reflects a *quality* assessment of outcome. As a result, if the outcome of a fire is bad, service output may be measured as low even if a lot of work was done by firefighters.

involve gas leakage or explosive materials. Also, arson is expected to result in a more severe outcome because these fires involve deliberate and destructive intentions.

Lastly, two types of density variables are included. First, the α term, *StationDensity*, represents the average service reach of fire stations in each city and functions as a proxy for how promptly an additional unit can arrive at the scene. Second, the δ term represents the population density of each city. This is a rough proxy for how likely it is for a fire to spread to another building, which is relevant to the dependent variable of this model. One possible critique may be that α and δ must be highly correlated because densely populated regions would also have a greater number of fire stations. However, a correlation of 0.45 suggests that this is not necessarily the case. Perhaps, other indicators of fire risk, such as the number of hazardous facilities, matter as much. Thus, we include both density variables in the model as they represent two different aspects of a city while not being highly correlated with each other.

5.2 Data Composition

Data were provided by the Fire Headquarters of Gyeonggi Province, South Korea. Thanks to the generous support from the Henry J. Copeland Independent Study Fund, the author was able to travel to South Korea, visit the Fire Headquarters, acquire the data set directly from the office of data management, and interview several administrators and firefighters. The original data set contains 98,000 observations and 144 variables. Each observation represents a fire incident that occurred between the years 2008 and 2018 in Gyeonggi Province. The variables contain detailed information about the fire incidents. All information was recorded in Korean and was manually translated into English by the author.

In 2014, the province implemented an electronic reporting system through which firefighters submitted information about their firefighting activity. Since then, firefighters were asked to click on a designated button on their electronic device (provided in every fire truck) when they arrived at the fire scene, finished suppressing the fire, and returned to the fire station. Thus, the exact time of these activities were electronically saved. Integral to this study is the exact duration of response to and suppression of fire. For this reason, all observations without this information, which are fires from before the year 2014, were omitted. Also, columns that contained overly detailed accounts of the fires in text format were excluded from consideration. As a result, the data set used in this empirical study consists of 49,000 observations and 80 variables.

The following sections provide a categorized list of all the columns and variables in the cleaned data set. Then, we will describe how the dependent variable for the service output model was constructed.

5.2.1 Administrative Information

When a fire occurs, the emergency call response center of Gyeonggi Province automatically designates a city fire department to an incident according to the location. Then, the closest available fire station is dispatched to the scene. Each fire incident is given a unique dispatch identification number. Also, jurisdictional information is recorded such as the name of the city fire department and station (or substation); and the name of the city and subregion in which the fire occurred. Names of the cities and subregions were encoded to numeric values for this study. The date and time of dispatch are included as well.

5.2.2 Distance and Time

Information about the distance from a fire scene to the city fire department, to the closest fire station, and to the closest substation (if any) is available. Using this information, a column indicating the minimum distance to the fire scene was constructed. The fire service closest to the fire scene is assumed to have been dispatched to the scene, which is a reasonable assumption because this is how the system automatically selects the fire service to be sent out. Also, response time is indicated in hours, minutes, and seconds. Using this information, a column indicating total response time in minutes was generated.

5.2.3 Suppression Activity

About 20 percent of the fires show to be suppressed before the firefighters' arrival, which is indicated by the binary variable, *by911*. These fire dispatches are not very meaningful in the discussion of firefighting power, but they still contain important information about response time and the placement decision of fire stations. However, in our analysis of service output in residential fires, these observations were dropped.

Suppression time indicates the duration of firefighting activity, starting at initial arrival and terminating at 'initial suppression'. When a fire is mostly suppressed and the situation is under control, firefighters report this to the headquarters. At this point, additional fire vehicles that were dispatched as back-ups return to their fire stations. Only the fire service in charge of the jurisdictional area remains and suppresses the rest of the fire. This 'final suppression' is not considered in our models.

Information about the number of firefighters, the number and types of fire trucks, and the involvement of other public services is also available. The types of fire vehicles are pumper, tank truck, ladder truck, chemical truck, rescue truck, ambulance, and helicopter. Other public services include the police, office of electricity and gas, army, and municipal civil servants. Oftentimes, volunteer firefighters are also mobilized in large fires, but all fires are responded to by full-time firefighters.

Fire Type	Count	Percent
Residential	10,953	22.09
Industrial	8,843	17.83
Vehicle	6,555	13.22
Service	4,127	8.32
Open fields	3,520	7.10
Commercial	2,586	5.22
Complex	394	0.79
Medical	359	0.72
School	318	0.64
Hazardous	26	0.05
Miscellaneous	11,903	24.02
Total	**49,586**	**100.00**

Table 5.5: Types of Fire by *areatype_legal*

5.2.4 Buildings and Objects

The variable *areatype_legal* consists of 15 types of objects or buildings where a fire occurred. A tabular summary is provided in Figure 5.5. Certain types of *buildings* are legally mandated to be equipped with fire prevention facilities. This requirement is indicated as a binary variable. The types of *object* that caught on fire exist as a column, which include flammable objects, furniture, electric devices, and several more.

Detailed information about the buildings is available. First, the area of a structure in km^2 is given. Second, the monetary value of real estate and insurance payments is recorded only in some observations. Third, there is information about the number of floors above and below ground in a building, where basement levels are indicated as negative integers. The floor on which a fire started is also indicated in the data set.

5.2.5 Cause of Fire

Binary variables indicating the cause of fire were considered. While not all of these causes may be indicated in the regression tables in the following chapter, the following variables were created and included in the service output model: arson, trash, kitchen, chemical, gas leak, and flammable objects.

5.2.6 Exogenous Conditions

Exogenous conditions that may affect response time are included in the cleaned data set, such as traffic and weather conditions. Using information about the time of dispatch, a binary variable indicating rush hour traffic was defined as times between 6−9 A.M. and 5−8 P.M. In the same vein, the seven days of a week were also considered to account for heavy traffic. Continuous variables such as temperature, humidity, and wind speed are present in the data set. In particular, binary variables for very high and very low temperatures were created.

5.2.7 Three Proxies for Outcome of a Fire

In order to measure the level of service output, a proxy must be chosen so that the abstract notion of *ex post* firefighting power can be quantified in an observable manner. Thus, it is inevitable that our measure of service output is chosen amongst different forms of **outcome** (consequence) of a fire. Recall that the ultimate goal of a fire service is to protect the lives and property of citizens in the case of a fire. To do so, the fire service aims to arrive at the fire scene as promptly as possible and stop the fire from spreading further. In this sense, the outcome of a fire incident can be considered in the following three aspects: human lives, property, and fire spread. We will assess if each of these aspects could be a measure of fire service output.

Human Lives

The data set contains information about the number of deaths and injuries due to a fire. Some observations also contain information about the number of persons who were evacuated or rescued by the fire services. Certainly, whether casualty could be a measure of service outcome is highly questionable. The argument of a proponent may be twofold. On one hand, one may argue that this variable can measure the value of a fire service because it is possible to measure the economic value of human life. For example, the value of a life lost may be measured in terms of the income that would have been generated had the person continued to live. However, the irreplaceability argument for human life (Singer, 1995) presents a strong case against this claim because individuals cannot be replaced by another being or a sum of money.

On the other hand, the proponent may argue that the number of lives saved can represent the value of the life-saving activity. After all, this number does not involve any quantitative evaluation of human life. However, one drawback is that areas with high population density will always have more lives saved; there are simply a larger number of people in these areas. Therefore, this measure fails to accurately capture the amount of work provided by firefighters.

For these two reasons, we conclude that neither the value nor the number of human lives can be an effective measure of service outcome.

Property

Columns related to property include: real estate value, real estate damaged by fire, movables value, and movables damaged by fire. The proportion of property saved in relation to the total property may be the most objective indicator of service outcome. This proportion would also be indicative of the 'outcome of interest to voters' (Duncombe and Yinger, 1993), which is the final form of service output.

Unfortunately, a great deal of measurement error was found in these variables. For example, some observations seem to be missing three or six digits of zeros. These errors were not systematic and thus could not be corrected. Information in other observations were unreliable because their total monetary value was less than the value of property destroyed. In short, variables relating to monetary value were practically impossible to be used in our models. Alternatively, the physical area of the original property and the area damaged in km^2 were considered as a proxy (Ignall, Rider, and Urbach, 1979).

Fire Spread

Also available in the data set is the information about the suppression point of a fire. For residential fires, the *suppression point* variable consists of five different entries, as shown in Table 5.6: object or room that caught on fire encoded as 1, entire floor as 2, other floors as 3, entire building as 4, and other buildings as 5. The variable indicates the specific location at which the residential fire was suppressed.

In order to construct a measure of the *severity* of fire spread, we needed

Code	Location of Fire
1	On object or in room
2	On entire floor
3	On other floors
4	On entire building
5	On other buildings

Table 5.6: Codes for Location of Fire (i, j)

a variable that would indicate the starting point of a fire based on a similar metric. Unfortunately, such column was not readily available in the data set. Thus, we chose to read through every single memo taken by the emergency call dispatcher and extract the relevant information. Almost all memos contained this information because citizens usually describe the fire when they make emergency calls. We were fortunate to receive support from the provincial fire headquarters not only because there were more than 7,000 residential fires, but also because they allowed us to remain consistent throughout the recording process. As a result, the *starting point* variable was constructed, which indicates the location of fire at the time when the emergency call was received. These entries were also encoded to integers ranging from 1 to 5.

5.3 Construction of the Dependent Variable

Following Jaldell (2005), this section will delineate our strategy of constructing the dependent variable for the service output model. This variable will consist of three ordinal levels that indicate the quality of outcome. To do so, we begin by establishing the notion of fire outcome.

5.3.1 Fire Outcomes

In previous sections, the phrase "outcome of a fire" was used as a general synonym for the word consequence. However, the term **fire outcome** hereafter will specifically refer to an indicator of fire spread. We claim that the degree of fire spread is a proxy for the performance of firefighters, as it is directly relevant to the *output* that is being produced. The degree of fire spread conveys information about how good or bad the consequences of a fire were, without taking into account the monetary value of the property. This was the argument presented in Section 3.2 for establishing fire spread as an objective measure of the amount of suppression activity produced by firefighters.

The variable *fire outcome* contains information about the location of a fire. A fire outcome is denoted v_{ij} where each of i and j are integers indicating the starting point and suppression point of a fire, as shown in Figure 5.1. For example, a fire outcome of v_{13} indicates that the fire was on the object or in the room when the emergency call was made; and that it was suppressed when it had spread to other floors. A total of fifteen fire outcomes can be considered, as shown in Table 5.3.1.

	Suppression Point (j)					
		1	**2**	**3**	**4**	**5**
Starting Point (i)	**1**	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}
	2		v_{22}	v_{23}	v_{24}	v_{25}
	3			v_{33}	v_{34}	v_{35}
	4				v_{44}	v_{45}
	5					v_{55}

Figure 5.1: Possible Fire Outcomes (v_{ij})

5.3.2 Ordering of Fire Outcomes

The next step is to order these fire outcomes into three levels using a set of ordering criteria. The primary criterion is the desirability of a fire outcome, which can be assessed through **inequality conditions** between fire outcomes. The secondary criterion is the level of **suppression effort** required to produce such a fire outcome. Deciding which fire outcome is "harder" to achieve involves subjective judgement. For this reason, we consulted firefighters and fire experts who have the hands-on experience of suppressing a fire.

Primary Criterion: Inequality Conditions

Jaldell (2005) recognizes that the fifteen fire outcomes from Figure 5.1 can be ordered both horizontally and vertically. On one hand, fire outcomes in each row can be ordered as follows:

$$v_{11} < v_{12} < v_{13} < v_{14} < v_{15}$$
$$v_{22} < v_{23} < v_{24} < v_{25}$$
$$v_{33} < v_{34} < v_{35}$$
$$v_{44} < v_{45}$$

where a "larger" fire outcome represents a more desirable one. For a given starting point of fire, a fire that spreads out further is less desirable. Put differently, for a given i, a larger j represents a worse outcome.

On the other hand, fire outcomes in each column can be ordered as follows:

$$v_{15} < v_{25} < v_{35} < v_{45} < v_{55}$$
$$v_{14} < v_{24} < v_{34} < v_{44}$$
$$v_{13} < v_{23} < v_{33}$$
$$v_{12} < v_{22}.$$

For a given suppression point of fire, a fire that ended up spreading less implies that the performance of firefighters was better. It is important to note that response time is a component of this performance because we have defined response time as an input for producing service output. For example, compare the fire outcomes v_{14} and v_{44}. The former indicates that the emergency call was made when the fire was on an object or in a room and that the fire ended up spreading to the entire building. In this case, the service provided was not effective or powerful enough to stop the fire from spreading. In the case of v_{44}, however, the fire was already very severe when the emergency call was made; it had spread to the entire building. The service provided was effective (due to a prompt response to fire) and powerful (due to an adequate number of firefighters and fire trucks) enough to prevent the fire from spreading out to another building. Thus, for given a j, a larger i represents a more desirable fire outcome.

Although the primary criterion allows for a mathematical and thus objective ordering of fire outcomes, the problem is that there remain combinations of fire outcomes for which such comparison is not possible. For example, there is no inequality condition available for v_{11} and v_{33}. In such cases, the secondary criterion was considered to order the fire outcomes.

Secondary Criterion: Suppression Effort

The secondary criterion for the ordering procedure is the level of suppression effort that is generally required to produce such an outcome. Since our dependent variable must reflect the amount of work produced by firefighters, it is reasonable to consider how hard it would have been to obtain a certain fire outcome in comparison to another.

In the case v_{11} and v_{33}, both are desirable outcomes in the sense that the fire was contained at the starting location. However, successfully preventing a fire from transferring to the entire building will probably require more effort than suppressing a fire on an object before it spreads out of the room. To make sure that these judgements aligned with the reality that firefighters actually confront, consultation of fire experts was conducted.

In applying the secondary criterion, different subgroups of residential fires were considered depending on the height of the building. Even if a set of two fires have the same fire outcome, the severity of the fire and the level of suppression effort may have been very different. For example, a fire outcome of v_{14} is likely to have been more destructive in a five-story building than in a single-story housing.

In the meantime, the floor on which the fire was located is also a crucial determinant of the suppression effort. For example, consider an identical set of two five-story buildings. A fire outcome of v_{22} would have involved a more powerful firefighting power if the fire started and was suppressed on the fifth floor, compared to a fire on the first floor. For this reason, the floor on fire was considered after the fires were divided into subgroups by the building height.

5.3.3 Grouping of Fire Outcomes

Finally, we can construct the dependent variable for the service output model. Based on the above ordering scheme, we now group the fire outcomes into three **levels**. The levels are ordinal in nature where $Y = 0$ represents the worst outcome and $Y = 2$ the best outcome. Fires with the worst outcome imply that the level of firefighting power produced by a fire service was low, whereas those with the best outcome indicate that a greater amount of suppression activity was produced by firefighters. As a result, we have established our dependent variable as an effective proxy for fire service output. The following section ordered and grouped the fire outcomes of a particular subgroup of residential buildings.

5.3.4 Example: A Subgroup of Buildings with 3^+ Stories

The subgroup consists of 2,496 fires that occurred on three-story buildings or higher, where the fire started on the second floor or higher. This subgroup was chosen as an example because all fifteen fire outcomes are theoretically plausi-

	Fire Outcomes					Count
$Y = 2$	v_{11} (1,929)	v_{22} (368)	v_{33} (18)	v_{44} (0)	v_{55} (0)	2,315
$Y = 1$	v_{12} (117)	v_{23} (32)				149
$Y = 0$	v_{13} (10)					21
	v_{14} (0)	v_{24} (2)	v_{34} (0)			
	v_{15} (0)	v_{25} (0)	v_{35} (2)	v_{45} (7)		

Figure 5.2: Ordering Example for Residential Buildings with 3^+ Stories

ble. In single-story houses, the location of fire cannot equal 3 because there are no "other floors" by definition. Meanwhile, the fact that the fire started on the second floor or higher will involve a nuanced judgement that is worth demonstrating in the body of this paper. The ordering procedure for other subgroups is provided in Appendix A, in the form of STATA output.

Figure 5.2 shows how the 2,496 fires were grouped into three levels. For visual purposes, the subscript *i*s representing the starting point were ordered horizontally, whereas *j*s representing the suppression point were ordered vertically. This grouping satisfies the first ordering criterion because it is consistent with the inequality conditions presented in Section 5.3.2. This can be checked by comparing the horizontal and vertical relationships.

In the meantime, since there are fifteen fire outcomes but only three levels, some fire outcomes end up being treated the same, qualitatively speaking. For example, even though it is clear that v_{13} is a better outcome than v_{15}, these two are placed in the same level of $Y = 0$. Our dependent variable does *not* differentiate between these two fire outcomes; v_{13} and v_{15} are "equally" bad. Although this lack of discretization can certainly be a limitation of our dependent variable, this is practically the best we can do with the given data set.

Another problem is that there are a disproportionately smaller number of severe fire outcomes. One possible solution could be to break down the best outcomes, $Y = 2$, into additional ordered levels. However, this distinction is likely become too arbitrary and subjective. Therefore, we chose to follow the three-level approach established by Jaldell (2005). While Jaldell (2019) also considers a binary dependent variable, we chose not to take this approach so that we can preserve as much information as possible.

The fire outcome v_{11} is categorized as the best outcome in the current example, whereas it is given a $Y = 1$ value in the subgroup for single-story buildings. Although putting out a fire before it spreads outside of the room is always desirable, the difficulty of doing so may vary according to circumstances. As far as *output* is concerned, we claim that a v_{11} on the third floor of a three-story building "deserves more credit" than a v_{11} on a single-story house.

Chapter 6

Results and Discussions

Following the model specifications presented in Section 4.1, three models were estimated. The models for **response time** and **suppression time** were estimated using the Ordinary Least Squares procedure. For these models, a set of analyses were conducted in order to check for the classical assumptions. Appendix B provides three plots for each model: a residual versus fitted plot, a normal probability plot, and a histogram of the residuals. The plots confirm the conditions relating to the error term. These analyses ensure that the OLS conditions were met and that we can rely on the results of our estimation.

Meanwhile, **service output** is estimated using an ordered probit model, the condition for which was briefly presented in Section 4.3.4. The proportional odds assumption, also called the parallel regression assumption, is practically impossible to check for. Therefore, we “assume” that the condition was met and proceed to present our estimation results.

For all three models, robustness tests were conducted to analyze the sensitivity of our results to different samples and selection of independent variables.

The results are presented in Appendix D.

6.1 Response Time Model

The response time model was estimated for 49,164 fire dispatches that occurred across the cities of Gyeonggi Province in the years between 2014 and 2018. The adjusted R^2 is 0.320, meaning that 32% of the variation in response time is explained by our regression equation.

Holding all else constant, an additional kilometer (or 0.621 mile) increase in **distance** leads to a 0.905 unit increase in response time, which is equivalent to 54.3 seconds. Considering that fire trucks typically travel at 40-80 kilometers (or 25-50 miles) per hour, the coefficient makes practical sense. A more meaningful analysis of the *Distance* variable will be provided at the end of this chapter.

Recall that *StationPopDen* refers to the population density per fire station. The interpretation of the coefficient is that having 1,000 more persons within a 1km^2 of land reduces response time by 0.068 minutes, or 4.08 seconds. To put this number into context, consider the two cities: Seoul, South Korea and Wooster, Ohio, U.S.A. Each city has a population density of 16,000/km^2 (or 42,000/mi^2) and 623/km^2 (or 1,613/mi^2) according to 2020 census data. For the sake of comparison, assume that our results can be generalized to other cities and countries. Then, the time it takes a fire truck in Seoul to travel an additional kilometer is expected to be shorter by 1.05 minutes compared to a fire truck in Wooster. Considering that an average fire in our data set has a distance of 3.6km, the difference in response times of the two cities can amount to nearly 4 minutes.

Table 6.1: Response Time Model (OLS)

Variable	**Coefficient**	**Std. Error**	**P-Value**
Distance	0.903***	0.032	0.000
StationPopDen	−0.068***	0.008	0.000
RushHour	0.371***	0.115	0.001
SnowRain	0.468***	0.071	0.000
Commercial	−0.752***	0.107	0.000
Residential	−0.151	0.105	0.153
School	−0.503**	0.213	0.018
Field	2.447***	0.189	0.000
Vehicle	−0.713***	0.129	0.000
Electrical	0.171***	0.042	0.000
Monday	0.197**	0.079	0.013
Friday	0.141*	0.082	0.086
Saturday	0.232**	0.086	0.007
Constant	4.027***	0.139	0.000
Obs. 49,164			
Adj. R^2 0.320			

Note: *p<0.1; **p<0.05; ***p<0.01

These results can also be understood in the context of population scale (Duncombe and Yinger, 1993). In regards with the intermediate output of interest, response time, we observe an *increasing* returns to population scale. That is, as population density increases, response time decreases, indicating that a fire station can serve more citizens and produce greater (intermediate) output in the form of shortened response time. Duncombe and Yinger (1993) present in their findings a *constant* returns to population scale. We are not sure if this is due to their lack of consideration for land size (in other words, the density aspect) or due to different output definitions. In their model, output is measured on the basis of monetary damage on property; the inverse of insurance payout is chosen as their proxy. In the present model, however, output is defined as response time in minutes.

Next, we move on to interpret the impact of various external conditions on response time. Snowy and rainy **weather** creates a half-minute delay in response time. Also, being in **rush hour** traffic, which we defined as 6-9 A.M. and 5-8 P.M, increases response time by 22.3 seconds. Mondays, Fridays, and Saturdays are associated with 8 to 14 seconds of prolonged travel times, compared to the Wednesday (base case). While not substantially large in magnitude, the coefficient estimates are all statistically significant at the 10% level or lower. These results can be useful when decision makers determine the times and days when bus-only lanes begin to be available for public use. In South Korea, like in many other urbanized places around the world, there are bus lanes designated to public transports. Citizen drivers are not allowed to drive on these lanes during certain hours of the day. In areas with particularly higher risks, the decision maker may designate certain days or hours and limit the use of these lanes for public transports and emergency vehicles.

We hypothesized that firefighters may respond to different **types** of fires with different levels of urgency or tension, perhaps unknowingly. Indeed, firefighters in our sample seem to have responded slightly more quickly to fires in commercial buildings, schools, and on vehicles. Fires on open fields pose a relatively low threat on the citizens and tend not to spread out severely. This may be an explanation to why these fires take 2.5 more minutes to be responded to, compared to miscellaneous fires (base case), controlling for all other variables including distance. While we expected a more prompt response to fires involving explosive or chemical materials, the coefficient estimates showed a positive sign and were not statistically significant.

6.2 Suppression Time Model

The suppression time model was estimated for 36,084 fires and is shown to explain 26.4% of the variation in suppression time. The sample size is smaller than the previous model by 13,137 because all fires that were either suppressed before the arrival of firefighters ($by911$=0) or had a suppression time of 0 ($SuppTime$=0) were dropped. Being an intermediate output, suppression time is not an accurate measure of the *quantity* of work performed by firefighters, compared to proxies such as fire spread or proportion of property damage. Despite this limitation, suppression time does provide useful information about the *duration* of service delivery, which is a key component of service output.

Recall that the **starting point**, i, takes one of the five integers between 1 and 5 which represents the location of fire. Fires that have already spread out from a room to the entire floor (i = 2) take 5.6 more minutes to suppress, compared to fires that were contained in the room (i = 1; base case). Fires that have already spread out to the entire building (i = 4) take about 8.5 more minutes to suppress, compared to the base case. The difference seems like an underestimation; using common sense, fires that are burning down an entire building should take much longer to suppress. This may be partially due to our sample containing various types and sizes of buildings, even though we

Table 6.2: Suppression Time Model (OLS)

Variable	Coefficient	Std. Error	P-Value
*StartPoint*2	5.572***	0.821	0.000
*StartPoint*3	4.994	3.258	0.220
*StartPoint*4	8.521***	3.165	0.007
RespTime	0.874***	0.175	0.000
Firefighter	0.243***	0.084	0.004
Pumper	3.212***	0.443	0.000
Tank	1.227***	0.462	0.008
Ladder	−0.673	0.873	0.441
Chemical	4.856***	0.909	0.000
RescueTruck	−3.061***	0.630	0.000
Ambulance	−2.062***	0.617	0.000
Helicopter	13.770**	6.063	0.023
ElectricGas	1.061***	0.291	0.001
Police	0.971***	0.291	0.000
CivilServant	0.347**	0.141	0.014
Hydrant	−0.398***	0.053	0.000
StationDensity	−0.032***	0.007	0.000
TempLow	7.131**	2.827	0.012
TempHigh	−1.719	1.115	0.123
ElecObject	−4.093***	0.474	0.000
Explosion	30.690***	4.830	0.000
Kitchen	−1.289**	0.650	0.047
Commercial	−3.725***	1.254	0.003
School	−5.483**	2.150)	0.011
Industrial	3.213**	1.509	0.033
Residential	−4.628***	1.181	0.000
Constant	−13.400***	2.553	0.000
Obs. 36,084			
Adj. R^2 0.264			

Note: *p<0.1; **p<0.05; ***p<0.01

did control for building types. When we estimate the same model on a sample of residential fires only, the coefficient estimates of *StartPoint*2 and *StartPoint*4 increase to 9 and 16 each, with p-values still close to zero.

Next, we examine the impact of all other intermediate outputs on suppression time. A prolonged **response time** will allow the fire to spread more, resulting in more suppression work to be performed. Our results show that an additional minute of delay in arrival leads to 52 more seconds of suppression time, holding all else constant.

Having one more **firefighter** results in a 15-second increase in suppression time. This is the *opposite* of what we expected. Our hypothesis was that having more firefighters would increase the firefighting power and result in shorter suppression time. Although this claim may be theoretically valid, our results remind us of the endogeneity concern that emerged during the model specification for dispatch level. In Section 4.1, we explained that a larger number of firefighters may be dispatched to the scene if the emergency call dispatcher decides that the fire is very severe or if the initial crew requests for additional, back-up crew. This is the strongest explanation that we can provide in response to the positive sign of the $Firefighter$ variable. Unfortunately, **pumper**, **tank**, and **chemical trucks** seem to suffer a similar endogeneity problem. The coefficient estimate on **helicopter** is particularly large, which is not very surprising given the fact that helicopters are typically dispatched in extremely severe fires. Similarly, personnel and vehicles from **other public services** produce statistically significant results that are opposite to our expectations.

In the meantime, **rescue trucks** and **ambulances** are shown to *reduce* suppression time, which is consistent with our expectation. Of course, the same argument could be made for these two variables as well; one may claim that larger fires would involve more victims and patients, resulting in higher dispatch level for these emergency vehicles. However, our results show that rescue trucks and ambulances shorten suppression time by approximately 3 and 2 minutes, respectively. The explanation would be that the presence of rescue and paramedic squads allows firefighters to focus on fire suppression only and be more productive in their firefighting activity.

The presence of one more fire **hydrant** per km^2 in a city is associated with a 24-second decrease in suppression time. This information may be useful when performing a cost-benefit analysis for installing additional fire hydrants in a city. The density of fire hydrants across the cities of Gyeonggi Province varies widely, ranging from 0.2 to 20 per km^2. Our results suggest that having 5 more hydrants in every 1km^2 of land can result in about a two-minute reduction in suppression time.

Recall that *StationDensity* represents the average **service reach** of a city's fire station:

$$StationDensity = \frac{area}{stations}.$$

This variable was included as a proxy for (the inverse of) density of the fire stations; if service reach is smaller, then fire stations are more densely located to each other, allowing additional units to arrive more promptly at fire scenes that are not within their jurisdiction. Our results show that a 1km^2 increase in service reach leads to a 0.032 unit increase in suppression time. To put this number into context, we can compare City 1 and City 3 from our data set. Each city has a service reach 5km^2 and 400km^2. This 395km^2 difference in service reach translates into a nearly 13 minutes longer suppression time in City 1. Fire stations in City 2 have the advantage of being located in closer proximity to each other, which can lead to faster fire suppression.

Temperature is available as a continuous variable in our data set. A few different functional forms were tested for the model but the results were not statistically significant. Instead, we categorized the variable into three groups: high, medium, and low. A temperature above 30°C (86°F) was defined as high; below -9°C (16°F); and anything in between as medium. As protective gears tend to trap heat, higher temperatures may reduce the productivity of firefighters. Meanwhile, low temperatures may create slippery surfaces and reduce the agility of firefighters. Our results show that lower temperature does have a negative effect on suppression activity, whereas the impact of higher temperature is statistically equivalent to zero.

Lastly, we interpret the results for the different **causes** and **types** of fires. All other things being equal, an industrial fire takes 3.2 more minutes to suppress. Fires that involve explosions take noticeably longer to suppress, requiring 30.6 more minutes than an average fire. Meanwhile, fires starting from the kitchen take less time to suppress compared to a 'miscellaneous' fire, which is the base case. Fires in commercial buildings, residential units, and schools also take less time to suppress. While these results are all statistically significant at the 5% level or lower, we are unable to provide a strong explanation and are left to conclude that further investigation is needed.

Generally speaking, shorter suppression time is a desirable outcome because it implies that the duration of threat posed by a fire was shortened. However, shorter suppression time does necessarily mean that a greater *amount* or better *quality* of work was performed. In fire incidents where initial severity is very high, suppression time may be longer even if firefighters did an excellent job. In these cases, it would be misleading to consider longer suppression time as an indication of bad outcome. Of course, if we had data on a control variable that "perfectly" captures the initial severity of fire, then suppression time would suffice as a measure of service output. Since this is not possible, we need another measure that reflects the nuanced differences between outcomes. We claim that **fire spread** can be such a measure because it is directly related to the quantity of suppression work. Another advantage of this variable as a proxy is that it is independent of the monetary value of the property and thus can be a more objective measure. Now, we proceed to estimate the service output model using the ordered dependent variable we constructed in Section 5.3.

6.3 Service Output Model

Table 6.3 shows the estimation results for the service output model. An ordered probit model was used due to the dependent variable being ordered and discrete in nature. Recall that the dependent variable consists of three categories, with $Y = 0$ indicating the worst outcome and $Y = 2$ the best outcome. As was explained in Section 5.3, the dependent variable is constructed in a way that captures both the *desirability* of a fire outcome based on spread level (primary criterion) and the level of suppression *effort* required to obtain such an outcome (secondary criterion).

It is crucial to note that the coefficient estimates should *not* be interpreted in the OLS manner. Practically speaking, one should only check the signs of

Table 6.3: Service Output Model (Ordered Probit)

Variable	Coeff	S.E.	P-Value	P.E.1	P.E.2	P.E.3
RespTime	-0.077^{***}	0.005	0.000	0.010	0.008	−0.018
SuppTime	-0.015^{***}	0.001	0.000	0.002	0.002	−0.004
Pumper	-3.212^{***}	0.019	0.000	0.024	0.021	−0.045
PumperSq	0.243^{***}	0.001	0.004	−0.001	−0.000	0.001
Tank	0.002^{***}	0.008	0.767	−0.000	−0.000	0.001
Ladder	0.673	0.047	0.000	−0.036	−0.031	0.067
Chemical	-4.856^{***}	0.040	0.003	0.015	0.013	−0.028
RescueTruck	3.061^{**}	0.027	0.000	−0.008	−0.007	0.016
Hydrant	0.021^{**}	0.009	0.032	−0.003	−0.002	0.005
Kitchen	0.823^{***}	0.108	0.000	−0.065	−0.070	0.136
Trash	0.271^{***}	0.083	0.001	−0.028	−0.027	0.056
Arson	0.177^{*}	0.102	0.083	−0.020	−0.018	0.037
Injury	0.068^{***}	0.021	0.001	−0.009	−0.007	0.016
Prevention	0.273^{***}	0.077	0.000	−0.030	−0.028	0.058
Saturday	0.235^{***}	0.070	0.001	−0.026	−0.024	0.050
*StationDensity*100	0.065^{**}	0.0000	0.037	−0.0001	−0.0001	0.0002
PopDensity	0.032^{**}	0.007	0.032	−0.0004	−0.0003	0.0008
Obs. 7,275						
Pseudo R^2 0.264						

Note: *p<0.1; **p<0.05; ***p<0.01

the coefficient estimates and their p-values. To observe the *partial* effect of an explanatory variable on the dependent variable, one must follow the procedure explained in Section 4.3.4 (Equations 4.34 to 4.36). We produced the average partial effects using the *mfx* command in STATA and presented the results in the last three columns of Table 6.3. The columns titled **P.E.1**, **P.E.2**, and **P.E.3** each indicates the partial effect of the explanatory variables on the number of $Y = 0$, $Y = 1$, and $Y = 2$ outcomes.

To demonstrate how the partial effects are interpreted, we take the *RespTime* variable as an example. A one-minute increase in **response time** results in a 1% increase in the number of $Y = 0$ outcomes, a 0.8% increase in the number of $Y = 1$ outcomes, and a 1.9% decrease in the number of $Y = 2$ outcomes. Simply put, an additional increase in response time results in *more* better outcomes and *fewer* worse outcomes. To put these figures into context, consider that the number of $Y = 0$, $Y = 1$, and $Y = 2$ outcomes in our data set is 1018, 629, and 5626, respectively. The partial effects suggest that a one-minute increase in response time is expected to result in 10 more $Y = 0$ outcomes, 5 more $Y = 1$ outcomes, and 107 less y=2 outcomes. In other words, we see a great deal of decrease in favorable fire outcomes and a bit of increase in severe fire outcomes.

Suppression time has a similar impact on service output but to a smaller degree. In fact, it makes logical sense that the partial effect of response time is greater. Longer response time implies a delay in the very commencement of the suppression activity. As a result, firefighters are faced with a more severe fire to start with. On the contrary, the impact of a prolonged suppression time will be spread across the entire duration of the suppression activity, the damage of which will be relatively weaker.

Unfortunately, the coefficient estimate on the *Pumper* variable shows to be negative again, just as we observed in the suppression time model. It would certainly be absurd to interpret this result in a literal sense and conclude that fire trucks cause worse fire outcomes. Rather, we suspect that this is due to the fact that severe fires that end up falling into the worst outcome category ($Y = 0$) will involve more firefighters and fire trucks. This argument has already been formulated when we specified the model for dispatch level in Section 5.1. It seems that the service output model suffers the same endogeneity problem that arose in the suppression time model.

The advantage of the model estimated by Jaldell (2005) is that a distinction is made between the number of **initial crew** and **additional crew**. In fact, the paper reports *positive* coefficient estimates on the two firefighter variables, *OwnFiremen* and *ExtraFiremen*, which is consistent with our theoretical understanding that increased firefighting power should lead to better fire outcomes. Our data set does not contain this information. It would have been very useful if our data set at least indicated the number of firefighters who arrived at the scene within, say, 6 minutes. This would have allowed us to decompose the contributions of the initial and the additional crew.

Given this limitation of our data, a series of robustness tests were carried out. One of our modest findings is that the coefficients on the fire truck variables do become positive when the sample is reduced to observations with *Pumper* < 4. Although the sample size drops to 1,031, we still confirm that an additional fire truck creates better outcome. We also find evidence for diminishing marginal returns to capital. The details of this sensitivity analysis are presented in Appendix D[1].

The results in Table 6.3 reinforce the common notion among fire experts and managers that the **rescue squad** significantly contributes to better fire outcomes. In South Korea, each city fire department has one rescue squad, consisting of individuals with demonstrated physical abilities. A recruitment track exclusive to former members of the special forces is certainly one factor that maintains the competence of the rescue squad. Also, distinguished firefighters with strong records on physical tests are often scouted to the rescue squad. Looking at our results, the high productivity of these workers seem to have substantial impact on the outcome of a fire. We conjecture that the *RescueTruck* variable is less susceptible to the endogeneity problem because they exist in limited numbers compared to pumper and tank trucks. This may have resulted in a positive sign of the coefficient estimate.

Meanwhile, increased accessibility to **fire hydrants** is shown to increase the number of better outcomes and reduce the number of worse outcomes. Recall that we observed the positive role of hydrants in reducing suppression time in the previous model, which seems to connect to producing better fire outcomes with less severe fire spread.

Many of our results contradict our theoretical understanding of the relationship between the variables. Increased mean **service reach** (multiplied by 1000 due to small decimal values) is associated with better fire outcomes. Specifically, an additional 1000km^2 of service reach results in 11 more $Y = 2$ outcomes

[1] Another method of sensitivity analysis was to consider different subsets of the data based on suppression time. That is, we estimated the service output model for fires with different ranges of suppression time. However, we were unable to find evidence to support the claim that more fire trucks lead to better outcomes.

and 1 less $Y = 0$ outcome. While the magnitude of the impact is practically negligible (considering that the average size of a *city* in Gyeonggi Province is 340km^2), the positive sign is still unexpected. It may be the case that the α_i terms are overly aggregated. Ideally, a better way to control for the density of the stations would be to compute the distance between fire stations and include it in the model. Likewise, the ideal alternative of the *PopDensity* variable would be disaggregated (individual) data on the distance between residential buildings. However, we did not have any information about these two variables.

We conclude this section with a brief discussion of our limitations and suggestions for data collection. Our initial motivation for estimating the service output model was to construct a **production function** for fire services, specifically their fire suppression activities. Unfortunately, we are only able to provide explanation for certain independent variables. Although we have attempted to interpret the coefficients on these variables to the best of our ability, the validity of our claims is threatened by the very fact that our interpretations hold true only for *certain* variables. To improve the service output model, it seems that having more information on the initial crew and additional crew will be necessary. The truck variables, which are the key factor inputs for fire service production, seems to suffer endogeneity problems. For future endeavors, we suggest that decomposing the contributions of the initial and additional crew will be necessary for establishing a more accurate and informative production function.

Another great strategy of improving the service output model would be to use a continuous dependent variable, which necessitates a more systematic collection of data. When we introduced the composition of our data in Section

5.2.7, we mentioned that there *were* data on the value of residential buildings and movable property, as well as the degree of monetary damage incurred on these two. However, measurement errors were too severe that these variables became practically useless; there were unexpectedly many observations where the original value of property was smaller than the monetary damage on the property, which was a sign of inaccurate information. In South Korea, there are designated officers whose responsibility is to investigate the cause and aftermath of large fire incidents. It may be useful to establish a more unified recording system so that measurement error becomes less prevalent in the data. This will allow one to construct a model that uses a *continuous* dependent variable, which will be greatly useful in that the method of analysis would not be restricted to logit or probit models. From a production theory standpoint, the availability of a continuous outcome variable will be particularly useful for analyzing the marginal contributions of each factor inputs and investigating the optimal level of dispatch level.

Chapter 7

Conclusions

The goal of our paper was twofold. First, we aimed to establish a coherent theory of fire service production that is capable of defining 'output'. As we discussed in the introduction chapter, economic studies have defined and measured output in many ways, resulting in a confusion as to what constitutes output in the case of fire services. By identifying the intermediate processes involved in producing fire suppression service, we have established the theoretical relationship between the inputs and the outputs. In particular, we defined dispatch level, response time, and suppression time as intermediate outputs. These outputs were then transformed into inputs to produce service output.

The second goal of the paper was to justify the status of response time as a factor of production. While there exist studies that treat response time as an input for producing fire services, the theoretical grounds were not sufficiently established. Our strategy was to identify the particular economic activity that 'produces' response time. To do so, we considered the production of fire services as a two-stage sequential optimization process. This allowed us to show that the second-stage decisions of the city fire departments result in the specific locations of fire stations; and that these locations are actualized into varied response times when fires occur. Now, we conclude this paper by summarizing the main points of each chapter, while also recognizing the limitations of our work and discussing further investigations that can be made.

In Chapter 1, we developed the novel idea of separating the *ex ante* and *ex post* production of fire services. We identified the two 'camps' of thoughts that differ over the definition of output. We introduced a way in which these two camps could be integrated into a coherent theoretical framework, the details of which followed in the next chapter.

In Chapter 2, the two stages of *ex ante* production were defined. The first-stage agent was the provincial government who distributes budget to its cities with the objective of attaining a desired level of social outcome. The second-stage agent was the city fire department who spatially allocates its fire stations with the goal of minimizing response time. As a result, varied stand-by levels and expected response times were produced. In an *ex post* framework, these outputs were each actualized into dispatch levels and response times. Under the influence of external factors, the duration of the suppression activity was determined. In our theoretical framework, all these factors eventually functioned as inputs for producing service output. That is, we established dispatch level, response time, and suppression time as factors of production.

In Chapter 3, we reviewed three papers that were relevant to our empirical models. Duncombe and Yinger (1993) provided an understanding of different returns to scale. In particular, the work revealed a connection between service reach and economies to population scale. Jaldell (2005) provided a conceptual framework that defined response time as a factor input. The work also established a statistical model of service output that defined response time as an input. Jaldell (2019) presented a detailed list of control variables that could improve the previous model. Extedning these two articles, we were able to show that response time is indeed a factor input, on both theoretical and statistical grounds.

In Chapter 4, we explained the statistical theory behind the three methods that were considered for this paper. The ordinary least squares method was introduced as an effective analytical tool that can derive the marginal contribution of an explanatory variable on the outcome variable. The probit model was shown to be useful for modeling discrete outcome variables. The instrumental variable approach emerged as a potential solution to our endogeneity problem, however, we were unable to identify the appropriate instruments due to insufficient data.

In Chapter 5, we presented our model specifications and described our data set. A total of four models were considered, the first three being the intermediate outputs from our theoretical framework and the last being the service output. Dispatch level, or the number of firefighters and fire trucks dispatched to a fire scene, was shown to be dependent upon the emergency call dispatcher's perception of the severity of a fire. Also, dispatch level was affected by the initial crew's request for additional back-up. We identified these factors as a potential source of endogeneity that may emerge in models involving dispatch level as an explanatory variable. Then, we specified the models for response time, suppression time, and service output as well. The rest of the chapter was dedicated to describing our data set in great detail.

In Chapter 6, we presented the estimation results of our regression models. We observed an increasing returns to population (density) scale when response time was defined as an output. Unfavorable weather and traffic conditions, as well as being on weekends, showed to prolong response time. We hypothesized that firefighters, perhaps unknowingly, may react more promptly to potentially severe fires. The results showed that fires on commercial, residential, and public buildings were responded 10-40 seconds more quickly than miscellaneous fires; travel times to field fires were 2.5 minutes longer compared to miscellaneous fires.

Meanwhile, we were cautious about reporting our results for the suppression time model. The reason was because dispatch level, particularly the primary responding units (firefighters, pumpers, and tank trucks), showed to increase suppression time. We did not interpret this result as firefighters and fire engines decreasing the efficiency of fire suppression; rather, we suspected that the endogeneity concern discussed earlier was affecting our results. In severe fires, a larger number of firefighters and fire trucks tend to be sent to the scene, to begin with. This point was elaborated in the model specification for the dispatch level.

With this limitation of the model in mind, we claimed that specialized vehicles such as the ladder truck, rescue truck, and ambulance contributed to reducing fire suppression time. Having a larger number of hydrants and fire stations in a given area also had a similar effect. We suggested that considering this 'density' aspect would be more appropriate (than considering the absolute quantity of resources) when using indicators or predictors that are specific to each *city*, rather than individual *fires*.

Finally, we presented our estimation results for the service output model. An ordered probit model was used, meaning that partial effects with respect to each variable had to be derived. The results suggested that the rescue squad plays an important role in producing better fire outcomes. This reinforced the understanding of fire administrators and managers that the skilled labor and physical qualifications of these members can have a strong, positive impact in fire incidents. The two time-oriented factors of production, response time and suppression time, showed to have a negative relationship with the severity of a fire outcome. We found this reasonable because prolonged response time and suppression time would allow the fire to spread out more and result in a worse fire outcome.

As was observed in the suppression time model, the key fire engines, pumpers and tank trucks, also showed to lead to worse fire outcomes in the service output model. We suspected that the congestion of fire trucks could be a reason, however, there was not enough evidence to support this claim. Although we could not fundamentally remedy the endogeneity problem, a set of robustness tests were conducted. In fires where less than 4 pumper trucks are dispatched, fire engines did show to contribute to better fire outcomes. This was consistent with our theoretical understanding that a larger number of firefighters and fire trucks would produce greater firefighting power. We concluded that an ideal solution to this endogeneity problem would be to consider the initial crew and additional crew separately. Unfortunately, this information was not available in our data set.

We conclude this paper by suggesting four directions in which this research can be extended. First, we have only considered fire suppression activities in this paper. Another important aspect of fire service provision is fire prevention. The theoretical framework should be expanded so that fire prevention efforts are considered as part of the service provision. Second, it may be useful to devise (or, practically speaking, collect data on) a continuous dependent variable. This will allow one to perform a more systematic and informative analysis of the factor inputs involved in the service output model. This will provide a more accurate decision tool for policy makers as well. Third, service output has been measured only in residential fires. In order to quantify the total output, one must also consider service output for fires in industrial buildings, public facilities, commercial spaces, forests, and so on. Lastly, it may be interesting to collect experimental data on how individual dispatchers respond to different emergency calls. With this information, the researcher may be able to mitigate the endogeneity problem that was the primary obstacle to the present study.

Appendix A

Ordering Fire Outcomes

Additional explanations will be added to the appendix upon completion of the oral examination for this thesis.

```
      name:  <unnamed>
       log:  /Users/alexhwang/Desktop/stata submit/hwang.smcl
  log type:  smcl
 opened on:  25 Mar 2020, 14:50:16

  1 . use "/Users/alexhwang/Desktop/stata submit/house_depVar.dta",clear
    (Written by R.              )

  2 .
  3 .                 **************************************************
  4 .                                 * Construction of output variable.
  5 .                 **************************************************
  6 .
  7 . * 1. Encode all spread outcomes.
  8 . generate str y= "."

  9 . replace y = "v11" if start_ordinal==1 & spread_ordinal==1
    variable y was str1 now str3
    (4,647 real changes made)

 10 . replace y = "v12" if start_ordinal==1 & spread_ordinal==2
    (571 real changes made)

 11 . replace y = "v13" if start_ordinal==1 & spread_ordinal==3
    (27 real changes made)

 12 . replace y = "v14" if start_ordinal==1 & spread_ordinal==4
    (172 real changes made)

 13 . replace y = "v15" if start_ordinal==1 & spread_ordinal==5
    (26 real changes made)

 14 . replace y = "v22" if start_ordinal==2 & spread_ordinal==2
    (920 real changes made)

 15 . replace y = "v23" if start_ordinal==2 & spread_ordinal==3
    (59 real changes made)

 16 . replace y = "v24" if start_ordinal==2 & spread_ordinal==4
    (542 real changes made)

 17 . replace y = "v25" if start_ordinal==2 & spread_ordinal==5
    (140 real changes made)

 18 . replace y = "v33" if start_ordinal==3 & spread_ordinal==3
    (30 real changes made)

 19 . replace y = "v34" if start_ordinal==3 & spread_ordinal==4
    (17 real changes made)

 20 . replace y = "v35" if start_ordinal==3 & spread_ordinal==5
    (11 real changes made)

 21 . replace y = "v44" if start_ordinal==4 & spread_ordinal==4
    (29 real changes made)

 22 . replace y = "v45" if start_ordinal==4 & spread_ordinal==5
    (83 real changes made)

 23 . replace y = "v55" if start_ordinal==5 & spread_ordinal==5
    (1 real change made)

 24 .
 25 . * 2. Order these outcomes into ordinal levels
 26 .
 27 . ***** 1-story buildings (2420 obs)
 28 . generate str y1 = "." if stories_aboveground ==1 | stories_aboveground ==0
    (4,871 missing values generated)

 29 . replace y1="0" if y=="v15" | y=="v25" | y=="v14"
    (338 real changes made)

 30 . replace y1="1" if y=="v11" | y=="v12" | y=="v24" | y=="v44" | y=="v45"
    (5,872 real changes made)

 31 . replace y1="2" if y=="v22"
    (920 real changes made)

 32 . gen y_levels = .
    (7,275 missing values generated)

 33 . replace y_levels = 0 if y1=="0"
    (338 real changes made)

 34 . replace y_levels = 1 if y1=="1"
    (5,872 real changes made)

 35 . replace y_levels = 2 if y1=="2"
    (920 real changes made)

 36 .
 37 . ***** 2-story buildings where floor_on_fire==1 (667 obs)
```

```
. generate str y2 = "." if stories_aboveground==2 & floor_on_fire==1
  (6,609 missing values generated)

. replace y2="0" if y=="v13" | y=="v14" | y=="v15" | y=="v24" | y=="v25" | y=="v35" | y=="v45"
  (1,001 real changes made)

. replace y2="1" if y=="v11" | y=="v12" | y=="v23" | y=="v34" | y=="v44"
  (5,323 real changes made)

. replace y2="2" if y=="v22" | y=="v33"
  (950 real changes made)

. replace y_levels = 0 if y2=="0"
  (663 real changes made)

. replace y_levels = 1 if y2=="1"
  (76 real changes made)

. replace y_levels = 2 if y2=="2"
  (30 real changes made)

.
. ***** 2-story buildings where floor_on_fire==2 (534 obs)
. generate str y3 = "." if stories_aboveground==2 & floor_on_fire==2
  (6,958 missing values generated)

. replace y3="0" if y=="v13" | y=="v14" | y=="v15" | y=="v24" | y=="v25" | y=="v35" | y=="v45"
  (1,001 real changes made)

. replace y3="1" if y=="v11" | y=="v12" | y=="v23" | y=="v34" | y=="v44"
  (5,323 real changes made)

. replace y3="2" if y=="v22" | y=="v33"
  (950 real changes made)

. replace y_levels = 0 if y3=="0"
  (0 real changes made)

. replace y_levels = 1 if y3=="1"
  (0 real changes made)

. replace y_levels = 2 if y3=="2"
  (0 real changes made)

.
. ***** 3+ story buildings where fire started on the first floor (879 obs)
. generate str y4 = "." if stories_aboveground !=0 & stories_aboveground !=1 & stories_aboveground !=2 & floor_on_fire==1
  (6,400 missing values generated)

. replace y4="0" if y=="v13" | y=="v14" | y=="v15" | y=="v24" | y=="v25" | y=="v34" | y=="v35" | y=="v45"
  (1,018 real changes made)

. replace y4="1" if y=="v11" | y=="v12" | y=="v23"
  (5,277 real changes made)

. replace y4="2" if y=="v22" | y=="v33" | y=="v44"
  (979 real changes made)

. replace y_levels = 0 if y4=="0"
  (17 real changes made)

. replace y_levels = 1 if y4=="1"
  (0 real changes made)

. replace y_levels = 2 if y4=="2"
  (29 real changes made)

.
. ***** 3+ story buildings where fire started on somewhere else than the first floor (2816 obs)
. generate str y5 = "." if stories_aboveground !=0 & stories_aboveground !=1 & stories_aboveground != 2 & floor_on_fire !=1
  (4,476 missing values generated)

. replace y5="0" if y=="v13" | y=="v14" | y=="v15" | y=="v24" | y=="v25" | y=="v34" | y=="v35" | y=="v45"
  (1,018 real changes made)

. replace y5="1" if y=="v12" | y=="v23"
  (630 real changes made)

. replace y5="2" if y=="v11" | y=="v22" | y=="v33" | y=="v44"
  (5,626 real changes made)

. replace y_levels = 0 if y5=="0"
  (0 real changes made)

. replace y_levels = 1 if y5=="1"
  (0 real changes made)

. replace y_levels = 2 if y5=="2"
  (4,647 real changes made)

.
. * Lastly, turn y_levels into labeled (ordinal).
. tostring y_levels, replace
```

```
   y_levels was float now str1

75 . encode y_levels, generate(y_level_final)

76 . drop y_levels

77 . rename y_level_final y_levels

78 . * My (final) dependent variable is called "y_levels".
79 .
80 . log close
      name:  <unnamed>
       log:  /Users/alexhwang/Desktop/stata submit/hwang.smcl
  log type:  smcl
 closed on:  25 Mar 2020, 14:50:16
```

Appendix B

Conditions for OLS

Additional explanations will be added to the appendix upon completion of the oral examination for this thesis.

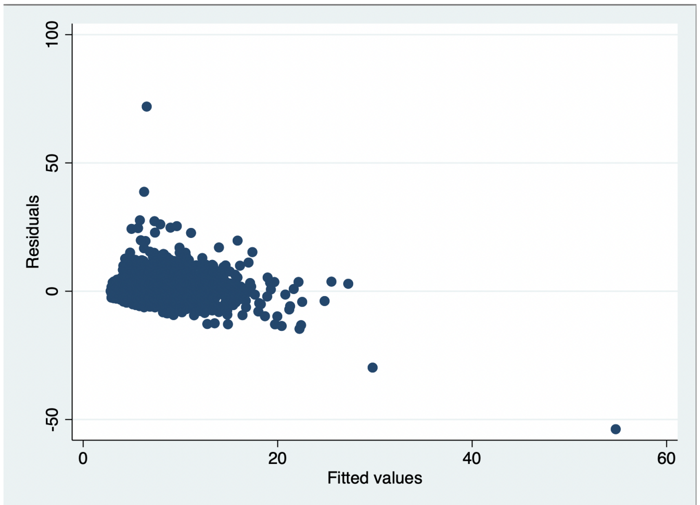

Figure B.1: Residual Plot for Response Time Model

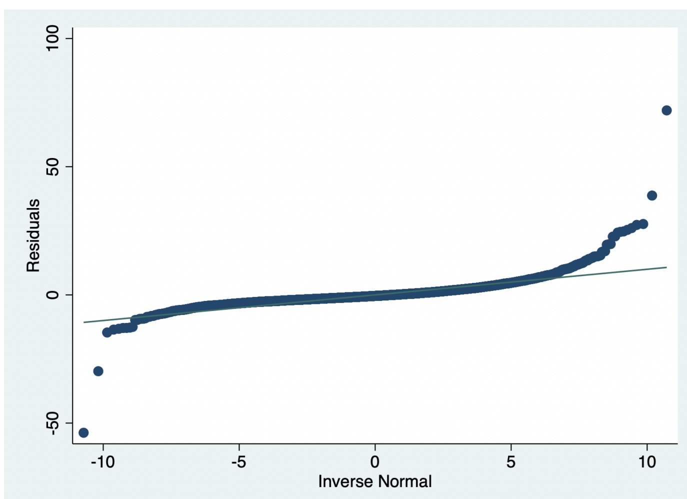

Figure B.2: Normality Plot for Response Time Model

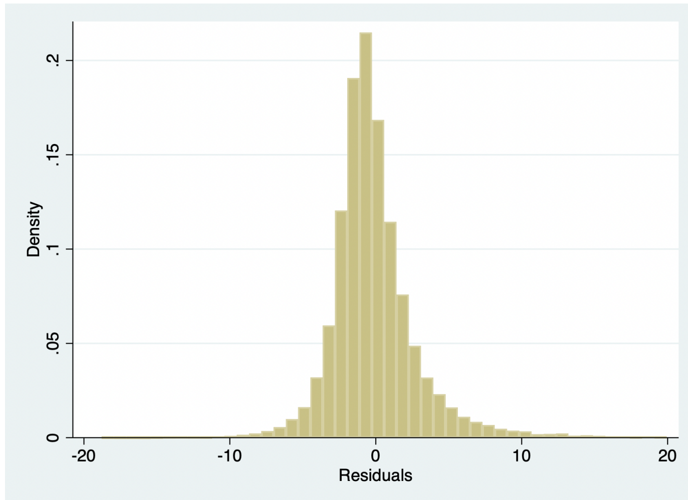

Figure B.3: Histogram of the Residuals for Response Time Model

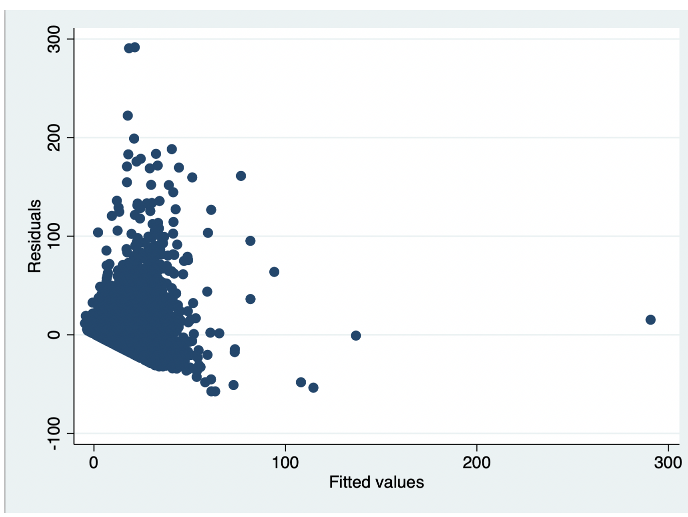

Figure B.4: Residual Plot for Suppression Time Model

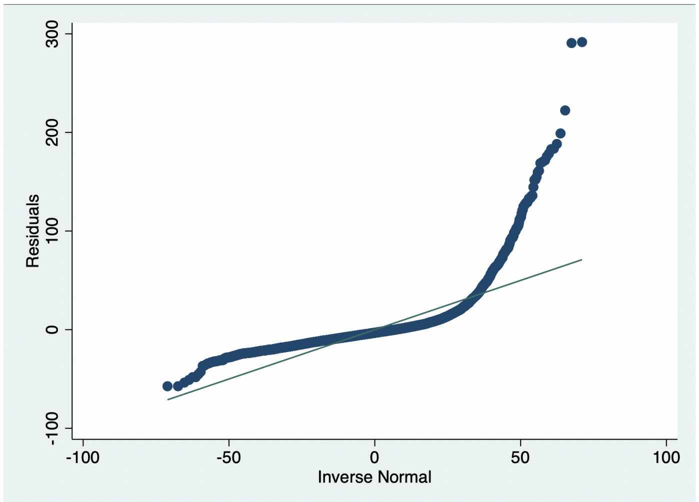

Figure B.5: Normality Plot for Suppression Time Model

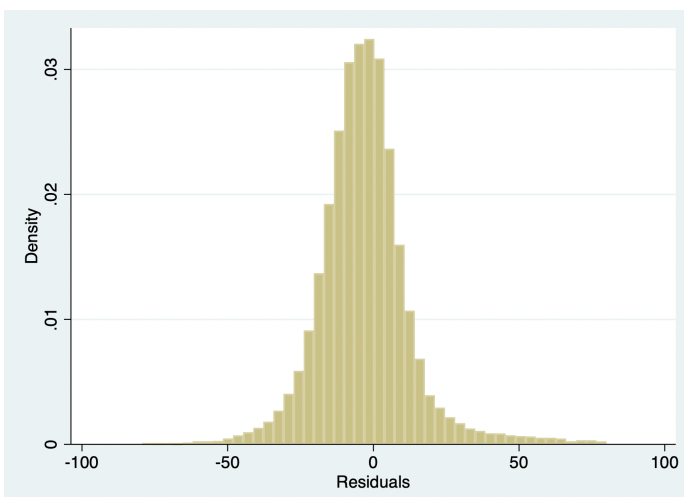

Figure B.6: Histogram of the Residuals for Suppression Time Model

Appendix C

Model Estimations

Additional explanations will be added to the appendix upon completion of the oral examination for this thesis.

```
      name:  <unnamed>
       log:  /Users/alexhwang/Desktop/stata submit/hwang_models.smcl
  log type:  smcl
 opened on:  25 Mar 2020, 15:29:49

1 . use "/Users/alexhwang/Desktop/stata submit/fire_models.dta", clear
  (Written by R.              )

2 .
3 . *** Model 1-a. Response Time
4 . reg RespTime Distance PopDensity rushhour snowrain commercial complex school hazard industrial medical field residential vehicle explosion miscellaneous service elec
  > _object rushhour##monday rushhour##tuesday rushhour##thursday rushhour##friday rushhour##saturday rushhour##sunday, r
  note: 1.rushhour omitted because of collinearity

  Linear regression                               Number of obs     =     49,164
                                                  F(29, 49134)      =     308.04
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.3205
                                                  Root MSE          =     4.2958
```

RespTime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Distance	.9031843	.0319153	28.30	0.000	.8406299	.9657387
PopDensity	-.0068329	.0007793	-8.77	0.000	-.0083605	-.0053054
rushhour	.3709858	.1145414	3.24	0.001	.1464833	.5954884
snowrain	.4679089	.0712413	6.57	0.000	.328275	.6075427
commercial	-.7518831	.1070792	-7.02	0.000	-.9617597	-.5420065
complex	-.0179504	.2181183	-0.08	0.934	-.4454649	.4095641
school	-.5032182	.2127122	-2.37	0.018	-.9201367	-.0862997
hazard	.0071402	.2778124	0.03	0.979	-.5373755	.551656
industrial	.1999755	.1144469	1.75	0.081	-.0243419	.4242929
medical	-.2591573	.3140498	-0.83	0.409	-.8746989	.3563842
field	2.447058	.1893476	12.92	0.000	2.075935	2.818182
residential	-.1506853	.1054753	-1.43	0.153	-.3574183	.0560477
vehicle	-.7128221	.1293631	-5.51	0.000	-.9663753	-.459269
explosion	.0755301	.1123284	0.67	0.501	-.144635	.2956952
miscellaneous	-.1221508	.1070804	-1.14	0.254	-.3320298	.0877281
service	-.7023661	.1044386	-6.73	0.000	-.9070672	-.4976651
elec_object	.1713523	.0421362	4.07	0.000	.0887649	.2539397
1.rushhour	0	(omitted)				
1.monday	.1970445	.0793677	2.48	0.013	.0414828	.3526062
rushhour#monday						
1 1	-.345954	.1527483	-2.26	0.024	-.6453426	-.0465655
1.tuesday	.0306151	.0762675	0.40	0.688	-.1188701	.1801003
rushhour#tuesday						
1 1	-.2104367	.1486843	-1.42	0.157	-.5018598	.0809863
1.thursday	.1044233	.0726533	1.44	0.151	-.0379781	.2468248
rushhour#thursday						
1 1	-.1232166	.2123072	-0.58	0.562	-.5393414	.2929082
1.friday	.1409641	.082141	1.72	0.086	-.0200334	.3019615
rushhour#friday						
1 1	-.1898676	.1592601	-1.19	0.233	-.5020194	.1222841
1.saturday	.2322812	.0856242	2.71	0.007	.0644567	.4001057
rushhour#saturday						
1 1	-.4997934	.1597272	-3.13	0.002	-.8128607	-.1867261
1.sunday	.0973212	.0829785	1.17	0.241	-.0653177	.25996
rushhour#sunday						
1 1	-.3640889	.1542229	-2.36	0.018	-.6663678	-.0618101
_cons	4.026786	.1386047	29.05	0.000	3.755119	4.298453

```
5 .
6 . *** Model 1-b. Response Time with city fixed effects
7 . reg RespTime Distance snowrain commercial complex school hazard industrial medical field residential vehicle explosion prevention elec_object i.city_code_label, r
  note: hazard omitted because of collinearity
  note: field omitted because of collinearity
  note: vehicle omitted because of collinearity

  Linear regression                               Number of obs     =     29,764
                                                  F(44, 29719)      =     180.03
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.3509
                                                  Root MSE          =     3.7157
```

RespTime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]

	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Distance	.9919508	.0699105	14.19	0.000	.8549231	1.128979
snowrain	.5038471	.0861885	5.85	0.000	.3349138	.6727804
commercial	-.200131	.0662828	-3.02	0.003	-.3300481	-.0702138
complex	.4409606	.1950235	2.26	0.024	.0587059	.8232152
school	.0755361	.1946027	0.39	0.698	-.3058937	.456966
hazard	0	(omitted)				
industrial	.5268836	.0986944	5.34	0.000	.3334383	.7203289
medical	.3108491	.3021664	1.03	0.304	-.2814103	.9031085
field	0	(omitted)				
residential	.3044721	.0615759	4.94	0.000	.1837806	.4251637
vehicle	0	(omitted)				
explosion	-.0346634	.1325123	-0.26	0.794	-.2943934	.2250666
prevention	-.3076673	.0555479	-5.54	0.000	-.4165436	-.1987911
elec_object	.1909326	.0488768	3.91	0.000	.095132	.2867331
city_code_label						
10	-1.21988	.311672	-3.91	0.000	-1.830771	-.6089889
11	-.4673277	.3485609	-1.34	0.180	-1.150522	.2158669
12	-.8269891	.3412308	-2.42	0.015	-1.495816	-.1581619
13	-.5491929	.3518796	-1.56	0.119	-1.238892	.1405066
14	-.3175408	.2955701	-1.07	0.283	-.8968712	.2617897
15	-.1134409	.3369012	-0.34	0.736	-.773782	.5469003
16	-.4979733	.2874219	-1.73	0.083	-1.061333	.0653862
17	-.7292897	.4017861	-1.82	0.070	-1.516808	.0582287
18	-.1983765	.2410585	-0.82	0.411	-.6708618	.2741088
19	-1.011266	.339746	-2.98	0.003	-1.677183	-.3453492
2	-.4934712	.3153935	-1.56	0.118	-1.111656	.1247139
20	.1501936	.3120488	0.48	0.630	-.4614357	.7618229
21	1.199462	.2794754	4.29	0.000	.6516777	1.747246
22	-.1224834	.2743556	-0.45	0.655	-.6602323	.4152656
23	-.4584465	.4272114	-1.07	0.283	-1.2958	.3789067
24	-.9733435	.3224857	-3.02	0.003	-1.60543	-.3412573
25	.3095022	.2753873	1.12	0.261	-.2302689	.8492734
26	-.5163569	.3472634	-1.49	0.137	-1.197008	.1642947
27	.1596242	.3598547	0.44	0.657	-.5457069	.8649553
28	-.4583297	.2728349	-1.68	0.093	-.9930981	.0764387
29	.0606186	.3293698	0.18	0.854	-.5849607	.7061979
3	-.3687803	.3874539	-0.95	0.341	-1.128207	.3906464
30	.1726597	.2774757	0.62	0.534	-.3712048	.7165241
31	-.6113588	.2908118	-2.10	0.036	-1.181363	-.0413549
32	-.4457077	.2772019	-1.61	0.108	-.9890357	.0976202
33	.1472117	.314485	0.47	0.640	-.4691928	.7636162
34	.793691	.2908945	2.73	0.006	.2235251	1.363857
4	-.7851696	.3377461	-2.32	0.020	-1.447167	-.1231724
5	-.0178839	.2612103	-0.07	0.945	-.5298674	.4940997
6	-.3478601	.342759	-1.01	0.310	-1.019683	.3239625
7	-.8667904	.3105256	-2.79	0.005	-1.475434	-.2581467
8	-.5970345	.2591726	-2.30	0.021	-1.105024	-.0890449
9	.4976161	.2842117	1.75	0.080	-.0594512	1.054684
_cons	3.467263	.4367753	7.94	0.000	2.611165	4.323362

```
.
.
. *** Model 2-a. Suppression Time
.
.         * Exlude suppression time=0 fires
.         drop if supptime==0
(8,859 observations deleted)

.         * Also exclude fires suppressed before arrival
.         keep if by911=="1"
(5,454 observations deleted)

.
. reg supptime start2 start3 start4 RespTime deptDensity_area firefighter electricgasofficers others_involved civilservant soldiers police pumper ladder tank chemical_
> truck rescue_truck ambulance helicopter HydrantDensity tempLow tempHigh humidity  injury  kitchen gasburn arson hazard_object elec_object trash explosion commercial
> complex school hazard industrial medical residential vehicle miscellaneous service, r

Linear regression                               Number of obs     =     34,908
                                                F(40, 34867)      =      77.91
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2630
                                                Root MSE          =     44.734
```

supptime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
start2	5.416935	.8355299	6.48	0.000	3.77927	7.054601
start3	2.045943	3.091035	0.66	0.508	-4.012583	8.10447
start4	8.953096	3.196576	2.80	0.005	2.687705	15.21849
RespTime	.8949863	.1782079	5.02	0.000	.545693	1.24428
deptDensity_area	.032836	.0074309	4.42	0.000	.0182711	.0474009
firefighter	.2418158	.0854409	2.83	0.005	.0743489	.4092827
electricgasofficers	.9651545	.2986178	3.23	0.001	.379854	1.550455
others_involved	.4297847	.231534	1.86	0.063	-.0240294	.8835987
civilservant	.3404994	.1409482	2.42	0.016	.0642363	.6167624
soldiers	.2095139	.1449894	1.45	0.148	-.0746699	.4936977
police	.9328326	.2101828	4.44	0.000	.5208676	1.344798
pumper	3.226054	.4563016	7.07	0.000	2.331689	4.12042
ladder	-.6544145	.8981294	-0.73	0.466	-2.414777	1.105948

tank	1.285948	.481062	2.67	0.008	.3430507	2.228845
chemical_truck	4.80149	.9211685	5.21	0.000	2.99597	6.60701
rescue_truck	-2.994108	.6407477	-4.67	0.000	-4.249994	-1.738222
ambulance	-2.114239	.5239188	-4.04	0.000	-3.141137	-1.087341
helicopter	13.75188	6.162989	2.23	0.026	1.672222	25.83154
HydrantDensity	-.3904553	.0545357	-7.16	0.000	-.4973471	-.2835635
tempLow	7.116602	2.866705	2.48	0.013	1.497768	12.73544
tempHigh	-2.133973	1.08417	-1.97	0.049	-4.25898	-.0089659
humidity	.0133011	.0061213	2.17	0.030	.0013032	.025299
injury	-2.105862	.3395531	-6.20	0.000	-2.771397	-1.440327
kitchen	-1.448985	.6609029	-2.19	0.028	-2.744376	-.1535942
gasburn	-14.14075	4.146888	-3.41	0.001	-22.26879	-6.012719
arson	-2.528229	.8914891	-2.84	0.005	-4.275576	-.7808819
hazard_object	-7.454278	1.822688	-4.09	0.000	-11.02681	-3.881751
elec_object	-4.101868	.4788073	-8.57	0.000	-5.040346	-3.16339
trash	-1.011262	.9342487	-1.08	0.279	-2.842419	.8198957
explosion	30.87032	4.889384	6.31	0.000	21.28697	40.45367
commercial	-3.794831	1.278958	-2.97	0.003	-6.301629	-1.288032
complex	-.8229492	2.367336	-0.35	0.728	-5.463004	3.817106
school	-5.554506	2.200598	-2.52	0.012	-9.867748	-1.241263
hazard	-12.34503	9.887194	-1.25	0.212	-31.72425	7.034188
industrial	3.086577	1.552524	1.99	0.047	.0435789	6.129574
medical	-2.386058	1.720666	-1.39	0.166	-5.758619	.9865022
residential	-4.542727	1.204607	-3.77	0.000	-6.903795	-2.181659
vehicle	-3.765483	1.355098	-2.78	0.005	-6.421519	-1.109446
miscellaneous	-.6249189	1.299105	-0.48	0.630	-3.171206	1.921368
service	-3.463337	1.25075	-2.77	0.006	-5.914846	-1.011827
_cons	-13.7478	2.611151	-5.27	0.000	-18.86574	-8.629864

```
18 .
19 . *** Model 2-b. Suppression Time for only residential fires
20 .
21 .         * Drop all other fires but residential
22 .         keep if areatype_legal=="Residential"
   (28,158 observations deleted)

23 .
24 . reg supptime start2 start3 start4  RespTime deptDensity_area firefighter electricgasofficers others_involved civilservant soldiers police pumper ladder tank chemical
   > _truck rescue_truck ambulance helicopter Hydrants tempLow tempHigh humidity snowrain injury death kitchen gasburn arson hazard_object elec_object trash explosion, r

   Linear regression                               Number of obs     =      6,757
                                                   F(32, 6724)       =     348.31
                                                   Prob > F          =     0.0000
                                                   R-squared         =     0.2311
                                                   Root MSE          =      19.65
```

supptime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
start2	8.973703	.7049928	12.73	0.000	7.591694	10.35571
start3	8.200432	2.569119	3.19	0.001	3.164145	13.23672
start4	16.20118	2.870155	5.64	0.000	10.57477	21.8276
RespTime	.9458282	.0976369	9.69	0.000	.7544291	1.137227
deptDensity_area	.0308327	.0078016	3.95	0.000	.0155391	.0461263
firefighter	-.1532241	.0605392	-2.53	0.011	-.2719001	-.034548
electricgasofficers	1.215459	.4178301	2.91	0.004	.39638	2.034539
others_involved	.4838874	.3042921	1.59	0.112	-.1126215	1.080396
civilservant	1.335851	.6112037	2.19	0.029	.1376979	2.534004
soldiers	-.0139798	.0279496	-0.50	0.617	-.06877	.0408103
police	.2200836	.1471193	1.50	0.135	-.0683168	.508484
pumper	1.707872	.2226768	7.67	0.000	1.271355	2.144389
ladder	-2.536343	.4164104	-6.09	0.000	-3.352639	-1.720047
tank	.1252997	.1897851	0.66	0.509	-.2467393	.4973386
chemical_truck	2.053033	.6398636	3.21	0.001	.7986977	3.307369
rescue_truck	.1334819	.365722	0.36	0.715	-.5834491	.8504129
ambulance	-.4425119	.3191591	-1.39	0.166	-1.068165	.1831411
helicopter	11.11681	6.202597	1.79	0.073	-1.042247	23.27586
Hydrants	-.0033255	.0007365	-4.51	0.000	-.0047694	-.0018816
tempLow	4.660909	1.896559	2.46	0.014	.9430534	8.378765
tempHigh	-1.738088	.8593324	-2.02	0.043	-3.422652	-.0535239
humidity	.0188535	.0080716	2.34	0.020	.0030307	.0346764
snowrain	.0013039	.8336291	0.00	0.999	-1.632873	1.635481
injury	.707843	.3016417	2.35	0.019	.1165297	1.299156
death	5.129758	2.403806	2.13	0.033	.4175367	9.84198
kitchen	-2.800691	.5344019	-5.24	0.000	-3.848288	-1.753094
gasburn	-3.944186	2.602864	-1.52	0.130	-9.046623	1.158252
arson	-2.645048	1.066657	-2.48	0.013	-4.736034	-.5540612
hazard_object	-6.264681	3.01434	-2.08	0.038	-12.17374	-.35562
elec_object	-2.785709	.6010867	-4.63	0.000	-3.96403	-1.607389
trash	-3.252211	.6781984	-4.80	0.000	-4.581694	-1.922727
explosion	-5.032721	1.642809	-3.06	0.002	-8.253148	-1.812295
_cons	.0941246	1.212606	0.08	0.938	-2.282967	2.471216

```
25 .
26 . * Open House Fire dataset again
27 . use "/Users/alexhwang/Desktop/stata submit/house_new2.dta", clear
   (Written by R.              )

28 .
29 . *** Model 3-a. Service Output
```

```
30 . oprobit y_levels RespTime supptime pumper pumper_sq tank ladder chemical_truck rescue_truck HydrantDensity kitchen trash arson injury prevention saturday deptDensity
   > _area PopDensity ambulance soldiers police volunteer electricgasofficers civilservant hazard_object elec_object   floor snowrain monday tuesday thursday friday satur
   > day sunday rushhour

   note: saturday omitted because of collinearity
   Iteration 0:   log likelihood = -4998.3127
   Iteration 1:   log likelihood = -3766.9936
   Iteration 2:   log likelihood = -3708.4261
   Iteration 3:   log likelihood = -3706.1305
   Iteration 4:   log likelihood = -3706.0639
   Iteration 5:   log likelihood = -3706.0636

   Ordered probit regression                       Number of obs     =      7,275
                                                   LR chi2(33)       =    2584.50
                                                   Prob > chi2       =     0.0000
   Log likelihood = -3706.0636                     Pseudo R2         =     0.2585

   ------------------------------------------------------------------------------------
              y_levels |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
   --------------------+---------------------------------------------------------------
              RespTime |  -.0769842   .0046463   -16.57   0.000    -.0860908   -.0678777
              supptime |  -.0150568    .000887   -16.98   0.000    -.0167952   -.0133183
                pumper |  -.1938398   .0186438   -10.40   0.000     -.230381   -.1572985
             pumper_sq |   .0041158   .0010153     4.05   0.000     .0021259    .0061057
                  tank |   .0024426   .0082485     0.30   0.767    -.0137242    .0186095
                ladder |   .2881778    .046534     6.19   0.000     .1969728    .3793827
        chemical_truck |  -.1196123   .0399569    -2.99   0.003    -.1979264   -.0412983
          rescue_truck |   .0641824   .0266296     2.41   0.016     .0119894    .1163755
        HydrantDensity |    .020941   .0097765     2.14   0.032     .0017793    .0401026
               kitchen |   .8228424   .1077427     7.64   0.000     .6116707    1.034014
                 trash |   .2711256    .083027     3.27   0.001     .1083956    .4338556
                 arson |   .1773681   .1023634     1.73   0.083    -.0232605    .3779966
                injury |    .067928   .0205491     3.31   0.001     .0276524    .1082035
            prevention |   .2726915   .0770438     3.54   0.000     .1216884    .4236947
              saturday |   .2349427   .0696284     3.37   0.001     .0984735    .3714119
      deptDensity_area |   .0006515   .0003125     2.08   0.037      .000039    .0012639
            PopDensity |   .0032227   .0015034     2.14   0.032     .0002761    .0061693
             ambulance |  -.0043295    .025805    -0.17   0.867    -.0549065    .0462474
              soldiers |   .0025181   .0022954     1.10   0.273    -.0019808    .0070171
                police |  -.0473147   .0075252    -6.29   0.000    -.0620638   -.0325656
        volunteer_fire |   -.024595   .0043869    -5.61   0.000    -.0331931   -.0159968
   electricgasofficers |  -.0336012   .0113584    -2.96   0.003    -.0558634   -.0113391
          civilservant |  -.0361656   .0170637    -2.12   0.034      -.06961   -.0027213
         hazard_object |   .3702313   .1927183     1.92   0.055    -.0074896    .7479523
           elec_object |   .0232451    .043724     0.53   0.595    -.0624524    .1089426
                 floor |   .0241839   .0076275     3.17   0.002     .0092342    .0391336
              snowrain |   .1015896    .059516     1.71   0.088    -.0150596    .2182388
                monday |   .0345423   .0686426     0.50   0.615    -.0999948    .1690793
               tuesday |   .0657122   .0688488     0.95   0.340    -.0692289    .2006533
              thursday |   .1121403   .0697473     1.61   0.108     -.024562    .2488426
                friday |   .0368272   .0671115     0.55   0.583    -.0947088    .1683633
              saturday |          0  (omitted)
                sunday |    .134053   .0692244     1.94   0.053    -.0016243    .2697304
              rushhour |   .0367414   .0433275     0.85   0.396     -.048179    .1216617
   --------------------+---------------------------------------------------------------
                 /cut1 |  -7.308688   .4566901                     -8.203785   -6.413592
                 /cut2 |  -2.617288   .1099216                     -2.832731   -2.401846
                 /cut3 |  -2.132827   .1084331                     -2.345352   -1.920302
   ------------------------------------------------------------------------------------

31 .
32 . * Marginal effects
33 . mfx, predict(outcome(2))

   Marginal effects after oprobit
         y  = Pr(y_levels==2) (predict, outcome(2))
            =  .06397641
   ------------------------------------------------------------------------------
   variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
   ---------+--------------------------------------------------------------------
   RespTime |   .0096415      .00066   14.71   0.000   .008357  .010926   6.56964
   supptime |   .0018857      .00013   14.16   0.000   .001625  .002147   14.6121
     pumper |   .0242765      .00237   10.23   0.000   .019626  .028927    5.8521
   pumper~q |  -.0005155      .00013   -4.07   0.000  -.000763 -.000267   40.0231
       tank |  -.0003059      .00103   -0.30   0.767  -.002331  .001719   3.20619
     ladder |  -.0360914      .00593   -6.09   0.000  -.047711 -.024472   .327148
   chemic~k |   .0149803      .00503    2.98   0.003   .005131  .024829   .188866
   rescue~k |  -.0080382      .00334   -2.40   0.016  -.014591 -.001485   1.30502
   Hydran~y |  -.0026226      .00123   -2.14   0.032  -.005024 -.000222   5.65955
   kitchen* |  -.0654812      .00492  -13.31   0.000  -.075121 -.055841   .125636
     trash* |  -.0283555      .00717   -3.96   0.000  -.042404 -.014307   .067491
     arson* |  -.0195889      .00989   -1.98   0.048  -.038974 -.000204   .039038
     injury |  -.0085073       .0026   -3.27   0.001  -.013608 -.003407   .122062
   preven~n* |   -.029941      .00736   -4.07   0.000  -.044358 -.015524   .180344
   saturday* |  -.0259317      .00675   -3.84   0.000  -.039168 -.012695   .146254
   deptDe~a |  -.0000816      .00004   -2.08   0.037  -.000158 -4.8e-06   49.0174
   PopDen~y |  -.0004036      .00019   -2.14   0.032  -.000773 -.000035   38.3167
   ambula~e |   .0005422      .00323    0.17   0.867   -.00579  .006875   1.22804
   soldiers |  -.0003154      .00029   -1.10   0.273  -.000879  .000248   .154777
     police |   .0059257      .00095    6.24   0.000   .004064  .007788   3.20962
   volunt~e |   .0030803      .00057    5.45   0.000   .001973  .004188   1.01718
   electr~s |   .0042082      .00143    2.95   0.003   .001409  .007008   .712852
   civils~t |   .0045294      .00214    2.11   0.035   .000327  .008732     .1189
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
hazard~t*	-.0349707	.01308	-2.67	0.008	-.060608	-.009334	.009622
elec_o~t*	-.0028847	.00538	-0.54	0.592	-.013424	.007654	.241237
floor	-.0030288	.00096	-3.17	0.002	-.004901	-.001157	2.93443
snowrain	-.0127231	.00746	-1.70	0.088	-.027349	.001903	.086735
monday*	-.0042442	.00827	-0.51	0.608	-.020459	.01197	.136907
tuesday*	-.007943	.00803	-0.99	0.322	-.023677	.007791	.145567
thursday*	-.0132115	.00772	-1.71	0.087	-.028345	.001922	.142131
friday*	-.0045227	.00808	-0.56	0.576	-.020358	.011313	.150241
sunday*	-.0156065	.00748	-2.09	0.037	-.030269	-.000944	.142405
rushhour*	-.0045362	.00527	-0.86	0.390	-.014872	.0058	.243849

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
34 . mfx, predict(outcome(3))
```

Marginal effects after oprobit
      y = Pr(y_levels==3) (predict, outcome(3))
         = .08571365

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
RespTime	.0082833	.00058	14.21	0.000	.007141	.009426	6.56964
supptime	.0016201	.00011	14.11	0.000	.001395	.001845	14.6121
pumper	.0208566	.00213	9.81	0.000	.016689	.025025	5.8521
pumper~q	-.0004429	.00011	-4.02	0.000	-.000659	-.000227	40.0231
tank	-.0002628	.00089	-0.30	0.767	-.002002	.001477	3.20619
ladder	-.0310071	.00514	-6.04	0.000	-.041076	-.020939	.327148
chemic~k	.01287	.00432	2.98	0.003	.004395	.021345	.188866
rescue~k	-.0069059	.00288	-2.40	0.016	-.012542	-.001269	1.30502
Hydran~y	-.0022532	.00105	-2.14	0.033	-.00432	-.000186	5.65955
kitchen*	-.0701319	.00654	-10.72	0.000	-.082958	-.057306	.125636
trash*	-.0272128	.0077	-3.53	0.000	-.042304	-.012122	.067491
arson*	-.0182355	.01	-1.82	0.068	-.037834	.001363	.039038
injury	-.0073089	.00223	-3.28	0.001	-.011681	-.002937	.122062
preven~n*	-.0278875	.00745	-3.74	0.000	-.042487	-.013288	.180344
saturday*	-.0241076	.0068	-3.55	0.000	-.037435	-.01078	.146254
deptDe~a	-.0000701	.00003	-2.08	0.038	-.000136	-4.0e-06	49.0174
PopDen~y	-.0003468	.00016	-2.14	0.032	-.000665	-.000029	38.3167
ambula~e	.0004658	.00278	0.17	0.867	-.004976	.005907	1.22804
soldiers	-.0002709	.00025	-1.10	0.273	-.000755	.000214	.154777
police	.0050909	.00083	6.15	0.000	.003469	.006713	3.20962
volunt~e	.0026464	.00048	5.47	0.000	.001697	.003595	1.01718
electr~s	.0036154	.00123	2.94	0.003	.001207	.006024	.712852
civils~t	.0038913	.00184	2.11	0.035	.00028	.007503	.1189
hazard~t*	-.0354553	.01586	-2.24	0.025	-.066535	-.004376	.009622
elec_o~t*	-.0024938	.00468	-0.53	0.594	-.011662	.006675	.241237
floor	-.0026021	.00083	-3.15	0.002	-.00422	-.000984	2.93443
snowrain	-.0109308	.00642	-1.70	0.088	-.023507	.001646	.086735
monday*	-.0036939	.0073	-0.51	0.613	-.017992	.010604	.136907
tuesday*	-.0069882	.00723	-0.97	0.334	-.021168	.007192	.145567
thursday*	-.0118158	.00719	-1.64	0.100	-.025914	.002282	.142131
friday*	-.0039376	.00713	-0.55	0.581	-.017912	.010037	.150241
sunday*	-.0140611	.00707	-1.99	0.047	-.027924	-.000198	.142405
rushhour*	-.0039351	.00462	-0.85	0.394	-.01299	.00512	.243849

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
35 . mfx, predict(outcome(4))
```

Marginal effects after oprobit
      y = Pr(y_levels==4) (predict, outcome(4))
         = .85030994

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
RespTime	-.0179248	.00112	-16.07	0.000	-.020111	-.015738	6.56964
supptime	-.0035058	.00022	-15.61	0.000	-.003946	-.003066	14.6121
pumper	-.0451331	.00429	-10.52	0.000	-.053545	-.036722	5.8521
pumper~q	.0009583	.00023	4.08	0.000	.000498	.001419	40.0231
tank	.0005687	.00192	0.30	0.767	-.003196	.004333	3.20619
ladder	.0670985	.01088	6.17	0.000	.045773	.088425	.327148
chemic~k	-.0278502	.00931	-2.99	0.003	-.046101	-.009599	.188866
rescue~k	.0149441	.0062	2.41	0.016	.002786	.027102	1.30502
Hydran~y	.0048758	.00227	2.14	0.032	.000417	.009334	5.65955
kitchen*	.1356131	.01066	12.72	0.000	.114713	.156514	.125636
trash*	.0555683	.01477	3.76	0.000	.02662	.084516	.067491
arson*	.0378243	.01986	1.90	0.057	-.001094	.076743	.039038
injury	.0158162	.00481	3.29	0.001	.00639	.025243	.122062
preven~n*	.0578285	.0147	3.93	0.000	.029016	.086641	.180344
saturday*	.0500393	.01347	3.72	0.000	.023642	.076436	.146254
deptDe~a	.0001517	.00007	2.08	0.037	9.1e-06	.000294	49.0174
PopDen~y	.0007504	.00035	2.15	0.032	.000065	.001436	38.3167
ambula~e	-.0010081	.00601	-0.17	0.867	-.012782	.010766	1.22804
soldiers	.0005863	.00053	1.10	0.273	-.000461	.001634	.154777
police	-.0110166	.00175	-6.31	0.000	-.01444	-.007593	3.20962
volunt~e	-.0057266	.00104	-5.53	0.000	-.007756	-.003698	1.01718
electr~s	-.0078236	.00265	-2.96	0.003	-.013011	-.002636	.712852
civils~t	-.0084207	.00398	-2.12	0.034	-.016219	-.000623	.1189
hazard~t*	.070426	.02885	2.44	0.015	.013872	.12698	.009622
elec_o~t*	.0053786	.01005	0.53	0.593	-.014326	.025084	.241237
floor	.0056309	.00177	3.18	0.001	.002157	.009105	2.93443
snowrain	.0236539	.01386	1.71	0.088	-.003513	.05082	.086735
monday*	.0079381	.01557	0.51	0.610	-.022571	.038447	.136907

```
    tuesday*     .0149312      .01526    0.98   0.328   -.01497  .044832   .145567
   thursday*     .0250273       .0149    1.68   0.093  -.004167  .054221   .142131
     friday*     .0084602      .01521    0.56   0.578  -.021345  .038266   .150241
     sunday*     .0296677      .01453    2.04   0.041   .001196  .058139   .142405
   rushhour*     .0084713      .00989    0.86   0.392  -.010913  .027856   .243849

(*) dy/dx is for discrete change of dummy variable from 0 to 1

36 .
37 . * Treat as OLS for correlation matrix analysis
38 . reg y_levels RespTime supptime firefighter firefighter_sq pumper pumper_sq tank tank_sq ladder ambulance floor_on_fire snowrain casualty_binary prevention kitchen ar
> son trash monday tuesday thursday friday saturday sunday rushhour,r

Linear regression                               Number of obs     =      7,275
                                                F(24, 7250)       =     103.56
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3157
                                                Root MSE          =     .59349
```

y_levels	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
RespTime	-.0409359	.003302	-12.40	0.000	-.0474088	-.0344629
supptime	-.0101294	.0006263	-16.17	0.000	-.0113571	-.0089016
firefighter	.0039661	.0016374	2.42	0.015	.0007562	.0071759
firefighter_sq	.0000143	4.04e-06	3.53	0.000	6.34e-06	.0000222
pumper	-.072183	.0080615	-8.95	0.000	-.0879859	-.05638
pumper_sq	.0002121	.000427	0.50	0.619	-.0006249	.0010491
tank	-.0159615	.0052096	-3.06	0.002	-.0261739	-.005749
tank_sq	.0000804	.0000231	3.48	0.001	.0000351	.0001257
ladder	.1356301	.0111497	12.16	0.000	.1137734	.1574867
ambulance	.0065463	.0100302	0.65	0.514	-.0131158	.0262084
floor_on_fire	.0047	.001275	3.69	0.000	.0022006	.0071995
snowrain	.0550383	.0216015	2.55	0.011	.0126931	.0973835
casualty_binary	-.1089281	.029735	-3.66	0.000	-.1672173	-.0506388
prevention	.0745619	.0135121	5.52	0.000	.0480742	.1010495
kitchen	.1434153	.0137624	10.42	0.000	.1164369	.1703936
arson	.1031693	.0325825	3.17	0.002	.0392981	.1670405
trash	.0987752	.023205	4.26	0.000	.0532867	.1442637
monday	-.0015022	.0280235	-0.05	0.957	-.0564364	.0534321
tuesday	.0242564	.0265485	0.91	0.361	-.0277864	.0762991
thursday	.0372375	.0264701	1.41	0.160	-.0146517	.0891267
friday	.0027624	.0268387	0.10	0.918	-.0498494	.0553742
saturday	.0741755	.0263239	2.82	0.005	.022573	.125778
sunday	.0434275	.0267246	1.62	0.104	-.0089606	.0958156
rushhour	.0071121	.0161192	0.44	0.659	-.0244863	.0387104
_cons	4.270405	.035947	118.80	0.000	4.199938	4.340871

```
39 .
40 . log close
      name:  <unnamed>
       log:  /Users/alexhwang/Desktop/stata submit/hwang_models.smcl
  log type:  smcl
 closed on:  25 Mar 2020, 15:30:27
```

Appendix D

Robustness Tests

Additional explanations will be added to the appendix upon completion of the oral examination for this thesis.

```
      name:  <unnamed>
       log:  /Users/alexhwang/Desktop/stata submit/robustness1.smcl
  log type:  smcl
 opened on:  25 Mar 2020, 15:36:15

1 .
2 . * Robustness test #1: Keeping only observations with Pumper < N
3 . forvalues x = 2(1)25{
  2. use "/Users/alexhwang/Desktop/Senior I.S./STATA/house_new2.dta", clear
  3. keep if pumper<`x'
  4. gen PopDensity = pop_density_city/100
  5. quietly oprobit y_levels pumper pumper_sq tank tank_sq RespTime supptime  ladder chemical_truck rescue_truck HydrantDensity kitchen trash arson injury prevention
> saturday deptDensity_area PopDensity ambulance soldiers police volunteer electricgasofficers civilservant   hazard_object elec_object   floor snowrain monday tuesday
>  thursday friday saturday sunday rushhour
  6. outreg2 using "/Users/alexhwang/Desktop/Robustness/pumper<`x'.doc", replace
  7. }
(Written by R.              )
(6,973 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<2.doc
dir : seeout
(Written by R.              )
(6,589 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<3.doc
dir : seeout
(Written by R.              )
(6,244 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<4.doc
dir : seeout
(Written by R.              )
(5,578 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<5.doc
dir : seeout
(Written by R.              )
(4,534 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<6.doc
dir : seeout
(Written by R.              )
(2,536 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<7.doc
dir : seeout
(Written by R.              )
(1,378 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<8.doc
dir : seeout
(Written by R.              )
(686 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<9.doc
dir : seeout
(Written by R.              )
(361 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<10.doc
dir : seeout
(Written by R.              )
(166 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<11.doc
dir : seeout
(Written by R.              )
(91 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<12.doc
dir : seeout
(Written by R.              )
(46 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<13.doc
dir : seeout
(Written by R.              )
(32 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<14.doc
dir : seeout
(Written by R.              )
(21 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<15.doc
dir : seeout
(Written by R.              )
(16 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<16.doc
dir : seeout
(Written by R.              )
(13 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<17.doc
dir : seeout
(Written by R.              )
(10 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<18.doc
dir : seeout
(Written by R.              )
(8 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<19.doc
dir : seeout
(Written by R.              )
(8 observations deleted)
/Users/alexhwang/Desktop/Robustness/pumper<20.doc
dir : seeout
```

```
  (Written by R.              )
  (6 observations deleted)
  /Users/alexhwang/Desktop/Robustness/pumper<21.doc
  dir : seeout
  (Written by R.              )
  (5 observations deleted)
  /Users/alexhwang/Desktop/Robustness/pumper<22.doc
  dir : seeout
  (Written by R.              )
  (3 observations deleted)
  /Users/alexhwang/Desktop/Robustness/pumper<23.doc
  dir : seeout
  (Written by R.              )
  (3 observations deleted)
  /Users/alexhwang/Desktop/Robustness/pumper<24.doc
  dir : seeout
  (Written by R.              )
  (3 observations deleted)
  /Users/alexhwang/Desktop/Robustness/pumper<25.doc
  dir : seeout

4 .
5 .
6 . * Robustness test #2: Treating observations with Pumpber > N the same
7 . forvalues x = 1(1)25{
  2. use "/Users/alexhwang/Desktop/Senior I.S./STATA/house_new2.dta", clear
  3. replace pumper=`x' if pumper>=`x'
  4. gen PopDensity = pop_density_city/100
  5. quietly oprobit y_levels pumper pumper_sq tank tank_sq RespTime supptime  ladder chemical_truck rescue_truck HydrantDensity kitchen trash arson injury prevention
> saturday deptDensity_area PopDensity ambulance soldiers police volunteer electricgasofficers civilservant   hazard_object elec_object   floor snowrain monday tuesday
>  thursday friday saturday sunday rushhour
  6. outreg2 using "/Users/alexhwang/Desktop/Robustness2/pumper=`x'.doc", replace
  7. }
  (Written by R.              )
  (6,973 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=1.doc
  dir : seeout
  (Written by R.              )
  (6,589 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=2.doc
  dir : seeout
  (Written by R.              )
  (6,244 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=3.doc
  dir : seeout
  (Written by R.              )
  (5,578 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=4.doc
  dir : seeout
  (Written by R.              )
  (4,534 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=5.doc
  dir : seeout
  (Written by R.              )
  (2,536 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=6.doc
  dir : seeout
  (Written by R.              )
  (1,378 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=7.doc
  dir : seeout
  (Written by R.              )
  (686 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=8.doc
  dir : seeout
  (Written by R.              )
  (361 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=9.doc
  dir : seeout
  (Written by R.              )
  (166 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=10.doc
  dir : seeout
  (Written by R.              )
  (91 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=11.doc
  dir : seeout
  (Written by R.              )
  (46 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=12.doc
  dir : seeout
  (Written by R.              )
  (32 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=13.doc
  dir : seeout
  (Written by R.              )
  (21 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=14.doc
  dir : seeout
  (Written by R.              )
  (16 real changes made)
  /Users/alexhwang/Desktop/Robustness2/pumper=15.doc
  dir : seeout
```

```
(Written by R.              )
(13 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=16.doc
dir : seeout
(Written by R.              )
(10 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=17.doc
dir : seeout
(Written by R.              )
(8 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=18.doc
dir : seeout
(Written by R.              )
(8 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=19.doc
dir : seeout
(Written by R.              )
(6 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=20.doc
dir : seeout
(Written by R.              )
(5 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=21.doc
dir : seeout
(Written by R.              )
(3 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=22.doc
dir : seeout
(Written by R.              )
(3 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=23.doc
dir : seeout
(Written by R.              )
(3 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=24.doc
dir : seeout
(Written by R.              )
(3 real changes made)
/Users/alexhwang/Desktop/Robustness2/pumper=25.doc
dir : seeout

8 .
9 . log close
       name:  <unnamed>
        log:  /Users/alexhwang/Desktop/stata submit/robustness1.smcl
   log type:  smcl
  closed on:  25 Mar 2020, 15:36:30
```

Bibliography

Ahlbrandt, R. C. (1973). Efficiency in the provision of fire services. *Public Choice*, 16:1–15.

Bradford, D. and Malt, W. O. (1969). The rising cost of local public services: some evidence and reflections. *National Tax Journal*, 22:185–202.

Christaller, W. and Baskin, C. W. (1966). *Central places in southern Germany*. Prentice-Hall.

Duncombe, W. D. (1992). Costs and factor substitution in the provision of local fire services. *The Review of Economics and Statistics*, 74:180–184.

Duncombe, W. D. and Yinger, J. (1993). An analysis of returns to scale in public production, with an application to fire protection. *Journal of Public Economics*, 52:49–72.

Greene, W. (2018). *Econometric analysis*. Pearson Education, Inc., 8th edition.

Hirsch, W. Z. (1973). *Urban economic analysis*. McGraw-Hill.

Ignall, E., Rider, K., and Urbach, R. (1979). Fire severity and response distance: initial findings. *Fire Technology*, 5:245–261.

Jaldell, H. (2005). Output specification and performance measurement: an ordinal output variable approach. *European Journal of Operational Research*, 16:525–535.

Jaldell, H. (2019). Measuring productive performance using binary and ordinal output variables: the case of the Swedish fire and rescue services. *International Journal of Production Research*, 57:907–917.

Jaldell, H., L. P. and Amornpetchsathaporn, A. (2014). Time is money, but how much? The monetary value of response time for Thai ambulance emergency services. *Value Health*, 17:555–560.

Kolesar, P. (1975). A model for predicting average fire company travel times. *Rand Institute Journal*, 1624:1–30.

Kristensen, O. (19). Public versus private provision of governmental services: the case of Danish fire protection services. *Urban Studies*, 20:1–9.

Perloff, J. M. (2016). *Theory and applications with calculus*. Pearson Education, 4th edition.

Singer, P. (1995). *Practical ethics*. Cambridge University Press, 3rd edition.

Studenmend, A. H. (2017). *Using Econometrics*. Pearson Education, Inc., 7th edition.

Taylor, B. M. (2015). Spatial modelling of emergency service response times. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 180.

Wallace, R. J. (1977). Productivity measurement in the fire service. *Public Productivity Review*, 2:12–36.